



Northwestern University Schizophrenia Data and Software Tool (NUSDAST)

Lei Wang^{1,2*}, Alex Kogan², Derin Cobia², Kathryn Alpert², Anthony Kolasny³, Michael I. Miller³ and Daniel Marcus^{4,5}

¹ Department of Radiology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

² Department of Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

³ Department of Biomedical Engineering, Center for Imaging Science, Johns Hopkins University, Baltimore, MD, USA

⁴ Department of Radiology, Washington University School of Medicine, St. Louis, MO, USA

⁵ Department of Psychology, Washington University School of Medicine, St. Louis, MO, USA

Edited by:

Maryann E. Martone, University of California, USA

Reviewed by:

Jeffrey S. Grethe, University of California, USA

Satrajit S. Ghosh, Massachusetts Institute of Technology, USA

*Correspondence:

Lei Wang, Department of Radiology and Psychiatry and Behavioral Sciences, Northwestern University Feinberg School of Medicine 710 N. Lake Shore Drive, 1322, Chicago, IL 60611, USA
e-mail: leiwang1@northwestern.edu

The schizophrenia research community has invested substantial resources on collecting, managing and sharing large neuroimaging datasets. As part of this effort, our group has collected high resolution magnetic resonance (MR) datasets from individuals with schizophrenia, their non-psychotic siblings, healthy controls and their siblings. This effort has resulted in a growing resource, the Northwestern University Schizophrenia Data and Software Tool (NUSDAST), an NIH-funded data sharing project to stimulate new research. This resource resides on XNAT Central, and it contains neuroimaging (MR scans, landmarks and surface maps for deep subcortical structures, and FreeSurfer cortical parcellation and measurement data), cognitive (cognitive domain scores for crystallized intelligence, working memory, episodic memory, and executive function), clinical (demographic, sibling relationship, SAPS and SANS psychopathology), and genetic (20 polymorphisms) data, collected from more than 450 subjects, most with 2-year longitudinal follow-up. A neuroimaging mapping, analysis and visualization software tool, CAWorks, is also part of this resource. Moreover, in making our existing neuroimaging data along with the associated meta-data and computational tools publically accessible, we have established a web-based information retrieval portal that allows the user to efficiently search the collection. This research-ready dataset meaningfully combines neuroimaging data with other relevant information, and it can be used to help facilitate advancing neuroimaging research. It is our hope that this effort will help to overcome some of the commonly recognized technical barriers in advancing neuroimaging research such as lack of local organization and standard descriptions.

Keywords: data sharing, neuroimaging, schizophrenia, computational anatomy, cognitive data, meta data

INTRODUCTION

Schizophrenia is a complex disease with heterogeneous clinical, behavioral, cognitive and genetic manifestations, and sharing of datasets is becoming essential in order to test hypotheses that can capture its variability and complexity (Poline et al., 2012). Case in point is the discovery of microRNA137 that succinctly illustrates the importance of data sharing: using computational biology techniques, Potkin et al. (2010) combined two previously published, separate datasets and discovered microRNA137 as a risk factor for schizophrenia. It should be noted that neither of the two distinct datasets had identified microRNA137. In a later confirmatory report on 51,695 individuals confirming microRNA137, the International Schizophrenia Consortium proclaimed that a new “cause” of schizophrenia had been found (Ripke et al., 2011).

In the neuroimaging front, the schizophrenia research community has witnessed consortium efforts such as the Functional Biomedical Informatics Research Network (FBIRN) (Friedman et al., 2008; Keator et al., 2008; Helmer et al., 2011), the Mind Clinical Imaging Consortium (MCIC) (Bockholt et al., 2010), and the Bipolar and Schizophrenia Network on Intermediate

Phenotypes (BSNIP) (Tammimga et al., 2013), which have warehoused or federated increasingly large magnetic resonance imaging (MRI) datasets. These data sharing efforts have led to consistent findings of brain network dysfunctions within the dorsolateral prefrontal cortex (Potkin et al., 2009) and the rest of the cerebral cortex (Potkin and Ford, 2009) during memory-related tasks (Kim et al., 2009) and at rest (Allen et al., 2011). In a recent study (Turner et al., 2012) that combined structural MRI data from Northwestern University (described in this paper), FBIRN, and MCIC, we used a multivariate independent components analysis and found that two components, one including the medial temporal lobe, insula, inferior frontal and temporal lobes, and the other including the posterior occipital lobe, showed significant familiarity.

In this paper, we describe the Northwestern University Schizophrenia Data and Software Tool (NUSDAST), an NIH-funded data sharing effort to stimulate new research. In making our existing neuroimaging data along with the associated meta-data and computational tools publically accessible, we have established a web-based information retrieval portal that allows

the user to efficiently search the collection. Search and retrieval of relevant research information identification is an important part of the data sharing process that justifies the effort spent on data collection and storage and makes it useful to other scientists.

With 451 subjects, the majority of whom with longitudinal data, NUSDAST is one of the largest single-site, single-platform neuroimaging datasets related to schizophrenia, making it a uniquely important resource to share with the research community. NUSDAST will benefit the neuroscience community in many ways. First, scientists will be able to use these data to generate or test new hypotheses related to abnormalities of brain structures and neural networks in individuals with schizophrenia. Second, scientists will be able to rapidly replicate findings produced using their own datasets. Third, the data could be used to test and validate new brain mapping tools. Further, automatic image processing pipelines designed for the analysis of these datasets, which consists of neuroanatomical landmarking and diffeomorphic mapping tools along with training and validation datasets, will enable others to study other MR datasets collected from other clinical samples.

DATA AVAILABLE FOR SHARING

The data presented here were collected through the support of two NIH-funded grants on schizophrenia: (1) Neuromorphometry in Schizophrenia (R01-MH056584), and (2) Conte Center for the Neuroscience of Mental Disorders (P50 MH071616). Through these projects, our group has collected high-resolution sMRI datasets from large cohorts of subjects using the same scanner platform and sequence protocols. We have also collected detailed clinical, cognitive and genetic information from these subjects. In addition, as part of the growing field of Computational Anatomy (CA) (Grenander and Miller, 1998; Csernansky et al., 2004a), our group has developed a suite of tools for mapping brain structures and characterizing schizophrenia-related abnormalities (Haller et al., 1997; Csernansky et al., 2004a; Wang et al., 2007b).

SUBJECTS

NUSDAST includes de-identified data in compliance with HIPAA¹ from 451 individuals with schizophrenia, their non-

¹De-identification consisted of stripping HIPAA-mandated identifiable information in research data (such as name, initials, and phone numbers, etc.). Procedures to further anonymize imaging data such as defacing were not performed in order to share the same imaging data that we used in our publications so that others can replicate our findings using their own algorithms if they so desire.

psychotic siblings, comparison subjects and their siblings. Longitudinal data also available on 171 subjects with schizophrenia (m/f = 114/57, age at baseline = 33.8 ± 12.5 years) and 170 controls (m/f = 86/84, age at baseline = 31.4 ± 13.8 years). Within these subjects, 18 subjects with schizophrenia and 30 controls returned for a second follow-up (i.e., 3 time points). The average (SD) follow-up interval was 2.19 (0.82) years for schizophrenia subjects and 2.28 (0.49) years for the controls. See **Table 1** below for baseline information.

Clinical data includes information based on specific criteria for clinical stability (Rastogi-Cruz and Csernansky, 1997) and clinical rating scales such as the Scale for the Assessment of Positive Symptoms (Andreasen, 1984) and Scale for the Assessment of Negative Symptoms (Andreasen, 1983) (see **Table 1** below for baseline information). Domains of psychopathology (i.e., psychotic symptoms, disorganized symptoms, and negative symptoms) (Andreasen et al., 1995) based on raw scales are also included. The reliability and practicality of using these scales in large populations of schizophrenic patients have been demonstrated by Andreasen et al. (1995). Symptom assessments were performed by personnel specially-trained for this purpose. Interrater reliability was monitored regularly for all rating scales, and rater training sessions, including the conjoint assessment of difficult cases, were held weekly. In these sessions, a variety of patients were interviewed in a group. Two established raters reached a consensus of item scores after the interview was completed, and then this “gold standard” score was compared with the rest of the group. New raters were trained by first participating in a minimum of six of these sessions. They were allowed to participate in ratings only after they had demonstrated satisfactory agreement with trained personnel.

MRI DATA

All MR scans were collected using the same 1.5 T Vision scanner platform (Siemens Medical Systems) at each time point. The Vision scanner had actively shielded gradients and echo-planar capability with very high gradient linearity (<0.4% over a 22-cm diameter spherical volume compared to 2–5% over 22-cm for our other scanners), which yielded anatomical images with virtually no distortion (<0.4% voxel displacement), critical to analyses of neuroanatomical structures. Using the same scanner provided stable longitudinal MR data throughout the entire period of data collection from 1998 to 2006.

Acquisition of all scans was performed at the Mallinckrodt Institute of Radiology at Washington University School of Medicine, where scanner stability (e.g., frequency, receiver gain,

Table 1 | Subject characteristics at baseline.

	Schizophrenia subjects	Control subjects	Schizophrenia siblings	Control siblings
N	171	170	44	66
Age at baseline (year)	33.8 (12.5 [17–63])	31.4 (13.8 [13–67])	N/A	N/A
Gender (male/female)	114/57	86/84	21/23	16/50
Race (Caucasian/African-American/Other)	90/78/3	61/105/2	17/27/0	16/50/0
Global SAPS score	11.1 [12.7 (0–81)]	0.06 [0.3 (0–4)]	0.5 [1.3 (0–7)]	N/A
Global SANS score	9.6 [10.7 (0–62)]	0.04 [1.7 (0–19)]	2.3 [4.8 (0–38)]	N/A

transmitter voltage, SNR) and artifacts were regularly monitored. Phantoms of known size were scanned to confirm image dimensions. Further tests and adjustments (shims, gradient calibrations, EPI switch delays, etc.) were made as needed. During each scan session, a small standardization object (i.e., vitamin-E gelcap) was placed on the left side of the forehead for each subject to clearly indicate laterality in the scans. Each scan session included a high-resolution T1 turbo-FLASH scan (Venkatesan and Haacke, 1997) and multiple (2–4) MPRAGE scans. Source MR scan data were in Siemens VISION IMA format and subsequently converted into Analyze™ format using in-house software. Since Analyze-format images may cause confusion with regard to laterality, even though the abovementioned vitamin-E gelcap information may help verify laterality, all Analyze-format images are being converted into NIFTI format and uploaded. The multiple MPRAGE images for each subject are aligned with the first image and averaged to create a low-noise image volume (Buckner et al., 2004). See **Table 2** for detailed scan protocol parameters.

NEUROIMAGING META-DATA

In our template-based brain mapping applications, we have focused on a network of structures previously implicated in the pathophysiology of schizophrenia (Weinberger et al., 1992; Csernansky and Bardgett, 1998; Goldman-Rakic, 1999). This network included regions with the prefrontal cortex (e.g., middle frontal gyrus—Brodmann area 46) (John et al., 2006; Harms et al., 2010), the cingulate gyrus (Qiu et al., 2007; Wang et al., 2007a), and the hippocampus (Wang et al., 2001; Csernansky et al., 2002), the parahippocampal gyrus (Karnik-Henry et al., 2012), as well as the thalamus (Csernansky et al., 2004a; Harms et al., 2007; Smith et al., 2011) and the basal ganglia (Mamah et al., 2007; Wang et al., 2008), which directly or indirectly link these structures via cortical-subcortical connections. We have constructed manual segmentation datasets for all these structures, which can be used for the validation of new computational methods. In addition, we have also used FreeSurfer (Desikan et al., 2006) to generate cortical surface parcellations and measures of cortical regional volume, thickness, and surface area (Cobia et al., 2011).

Template data

The templates for the hippocampus and amygdala were generated using a T1-weighted MR scan collected in a healthy subject. The templates for the thalamus and basal ganglia (caudate nucleus, putamen, nucleus accumbens and globus pallidus)

were generated using a seven-time averaged T1-weighted MR scan collected in another healthy subject. The segmentations were manually performed in these scans by consensus of experts using atlas guidelines (Duvernoy, 1988, 1991; Mai et al., 1997). Surfaces (byu format) of each structure were generated using the marching cubes algorithm (Lorensen and Cline, 1987; Claudio and Roberto, 1994). The left and right surfaces have corresponding nodes so that analyses of shape asymmetry can be performed.

Landmark and surface data

Mapping of the template MR scan occurred in a two-step process. First, it was coarsely aligned to each target scan using landmarks, and then the diffeomorphic map was applied. Surfaces for subcortical structures in the target scans were generated by carrying the template surfaces through these maps (Joshi et al., 1997; Csernansky et al., 2004b).

To facilitate our template-based mapping, global and local (i.e., structure-dependent) neuroanatomical landmarks were placed on the MR images. Landmark-based registration (Joshi et al., 1995) served to adjust the orientation and size for the head (based on global landmarks) and the subcortical structures of interest (based on local landmarks). *Global landmarks*: in each scan, twelve global landmarks were placed following procedures described in Haller et al. (1997): at the points where the anterior and posterior commissures intersected the midsagittal plane, and at the external boundaries of the cerebrum (anterior, posterior, superior, inferior and lateral). *Local landmarks*: (1) Hippocampus and amygdala were landmarked separately as follows. The most anterior and posterior boundaries of the structure were identified first and a line connecting these points created an anterior/posterior axis. Then in each of five equally distanced slices along this axis, four landmarks were placed at predetermined points in each slice. (2) Thalamus and basal ganglia were landmarked together as follows. The most anterior boundary of the caudate nucleus and the most posterior boundary of the thalamus were identified and a line connecting these points created an anterior/posterior axis. The region between the two points was then divided into five equally distanced slices along this axis and in each slice five landmarks were placed at predetermined places.

FreeSurfer data

All scans were processed through FreeSurfer Version 3.0.4 (Desikan et al., 2006) pipeline, with careful quality assurance as per FreeSurfer recommendations. All FreeSurfer data, including subcortical segmentation, cortical parcellation and surface, and regional measurement data have been made available.

COGNITIVE DATA

Schizophrenia subjects demonstrate a wide array of cognitive deficits (Gur et al., 2001). Data related to intellect, executive functioning (verbal and visual abstraction), and attention, as well as working and episodic memory are included in this data set. Measures of episodic memory included verbal and visual learning, and also the spontaneous and guided use of memory cues (Jacoby et al., 1993a,b, 2001; Jacoby, 1999). Our assessment

Table 2 | MR scan parameters.

Sequence	Protocol parameters
3D turbo-FLASH	TR = 20 ms, TE = 5.4 ms, flip = 30°, ACQ = 1, 256 × 256 matrix, 1 × 1 mm in-plane resolution, 180 slices, slice thickness 1 mm, 13:30 min scan time
3D MPRAGE (2–4 repeats)	TR = 9.7 ms, TE = 4 ms, flip = 10°, ACQ = 1, 256 × 256 matrix, 1 × 1 mm in-plane resolution, 128 slices, slice thickness 1.25 mm, 5:36 min scan time each

of working memory included maintenance and manipulation processes across both verbal and visual modalities, (Braver et al., 1997; Barch et al., 2002). To date, our schizophrenia subjects have demonstrated deficits across all predicted cognitive domains using this battery (Delawalla et al., 2006; Cobia et al., 2011).

At each visit, the subjects were administered a core battery of neuropsychological measures relevant to areas identified in prior studies of cognition in schizophrenia (Nuechterlein et al., 2004). Tasks were grouped into the following four domains:

Crystallized intelligence

Scaled scores from the Vocabulary subtest of Wechsler Adult Intelligence Scale [WMS-III; (Wechsler, 1997)].

Working memory

Scaled scores based on subtests from the Wechsler Memory Scale—Third Edition WMS-III; (Wechsler, 1997) including Digit Span (total forward and backwards), Spatial Span (total forward and backwards), and Letter-Number Sequencing, and also overall d-prime from the CPT-IP task (Cornblatt et al., 1988).

Episodic memory

Included scaled scores from the WMS-III Logical Memory and Family Pictures subtests.

Executive function

Included number of novel words generated on phonemic and semantic verbal fluency tasks (Benton, 1976; Benton et al., 1984), time to completion on the Trail Making Test Part B (Reitan and Wolfson, 1985), scaled scores on the WAIS-III Matrix Reasoning subtest, and number of perseverative errors on the Wisconsin Card Sorting Test (Heaton et al., 1993).

Cronbach's alpha (assessed in the standardization set of subjects) was 0.77, 0.78, and 0.70 for working memory, episodic memory, and executive function, respectively, in the individuals with schizophrenia, and 0.76, 0.65, and 0.67, respectively, in the control individuals.

GENOTYPING DATA

Blood for the isolation of DNA was collected from each of the subjects. These samples have been genotyped for a panel of 20 gene polymorphisms selected for their association with schizophrenia or their involvement in neurodevelopment. Examples of these genes include BDNF (rs6265), EGFR (rs10228436), FGF-2 (rs1048201), and IL-3 (rs40401). Morphometric measures (e.g., structural volume) of individuals are compared and contrasted with specific differences (i.e., Single Nucleotide Polymorphisms, or SNPs) in the genes of interest. These differences are qualified by testing whether or not each subject has a particular polymorphism, and then how many copies of that polymorphism they have. A subject can fall into one of three categories: both copies of the gene are polymorphism-free (homozygous), one copy is polymorphism-free whereas the other copy has the polymorphism (heterozygous), or both copies of the gene carry the polymorphism (homozygous).

Presently we have genotyping data on 117 subjects with schizophrenia and 58 controls. DNA samples in additional subjects are being analyzed. All available genotype data will be made available with the scans to users of the database.

DATA DICTIONARY

Along with the data, we provide a data dictionary of terms. In the dictionary, standard descriptions for which ontologies exist are used. We searched following sources for ontology: NeuroLEX (http://neurolex.org/wiki/Main_Page)—a semantic wiki for terms used in Neuroscience, the Neuroscience Information Framework (<http://www.neuinfo.org/>)—a dynamic resource of Web-based neuroscience data, materials, and tools (NeuroLEX terms are actually published in NIF), and NCI Metathesaurus (<http://ncimeta.nci.nih.gov/ncimbrowser/>)—a biomedical terminology database for translational and basic research. A detail list of these terms is presented in **Table 3**, and examples include “Socioeconomic Status,” “SAPS,” and “Cognitive Assessment.” For descriptions for which there are no standard ontologies, such as “Working Memory” or “Global Rating of Hallucinations,” we plan to work with NeuroLEX to arrive at standard definitions. The current version of the data dictionary can be downloaded through the data portal website, described below.

DATA SHARING MECHANISM

DATA SHARING ARCHITECTURE: XNAT AND XNAT CENTRAL

The collected MR datasets along with detailed clinical and cognitive information are archived using the eXtensible Neuroimaging Archive Toolkit (XNAT), an open source data management and productivity platform for biomedical imaging research. XNAT was developed by Neuroinformatics Research Group (NRG) at Washington University in St. Louis and the BIRN (Marcus et al., 2007a,b). It is widely used across the world and is a core component of the emerging NIH-backed biomedical informatics backbone, including the Biomedical Informatics Research Network (BIRN) and National Alliance for Medical Image Computing (NA-MIC). XNAT includes a secure database backend and a rich web-based user interface.

XNAT Central (<http://central.xnat.org/>) is a public access repository for neuroimaging and related data operated by the NRG. XNAT Central is built on the XNAT data management platform. XNAT Central includes a number of secure tools for storing and accessing images including a DICOM server, a web services API, and a user friendly website. Non-imaging data (e.g., demographics, clinical measures, derived measures) can be uploaded via spreadsheets or online forms. The website allows users to easily search for data across the various imaging and non-imaging measures, to visualize raw and processed images, and to download selected datasets.

XNAT CUSTOM SCHEMAS

XNAT relies on Extensible Markup Language (XML) schema documents to define the type of data that can be stored in the system. XML is the standard language for defining open and extensible data formats. XML format provides a number of benefits for data organization: it provides a way to uniformly describe data and

Table 3 | Table of available ontology for data dictionary terms.

Ontology	NeuroLex name and ID	NIF standard ontology ID	NCI methathesaurus ID
TERM			
Gender	Gender assessment birnlex_3026	nif_inv:birnlex_3026	C44177
Race	Race assessment birnlex_3040	nif_inv:birnlex_3040	C17049
Group	Control role birnlex_11017 ^a		
Ethnicity	Ethnicity assessment birnlex_3015	nif_inv:birnlex_3015	C16564
Marital status	Marital status assessment birnlex_3031	nif_inv:birnlex_3031	C25188
Type of housing/living arrangement	Living arrangement assessment birnlex_3030	nif_inv:birnlex_3030	C94852
Number of siblings			C102469
Number of children	Offspring cardinality assessment birnlex_3035	nif_inv:birnlex_3035	
Current occupation/job title	Occupation assessment birnlex_3036 ^b	nif_inv:birnlex_3036	C25193
Principle occupation	Occupation assessment birnlex_3036 ^b	nif_inv:birnlex_3036	C25193
Level of education	Education assessment birnlex_3014 ^c	nif_inv:birnlex_3014	
Years of schooling	Education assessment birnlex_3014 ^c	nif_inv:birnlex_3014	C17953
Father's level of education	Father's education birnlex_3021 ^d	nif_inv:birnlex_3021	
Father's years of schooling	Father's education birnlex_3021 ^d	nif_inv:birnlex_3021	
Mother's level of education	Mother's education birnlex_3023 ^e	nif_inv:birnlex_3023	
Mother's years of schooling	Mother's education birnlex_3023 ^e	nif_inv:birnlex_3023	
Socioeconomic status	Socio-Economic status birnlex_3048	nif_inv:birnlex_3048	C17468
Handedness			(CUI) C0023114
Edinburg handedness assessment	Edinburg handedness assessment birnlex_3013	nif_inv:birnlex_3013	
SAPS	Scale for the Assessment of Positive Symptoms birnlex_3045 ^f	nif_inv:birnlex_3045	
SANS	Scale for the Assessment of Negative Symptoms birnlex_3041 ^g	nif_inv:birnlex_3041	
Cognitive assessment	birnlex_2021	nif_inv:birnlex_2021	C0870300

^a Currently a proxy class to be replaced by its OBI (The Ontology for Biomedical Investigations) equivalent.

^b NeuroLex Occupation assessment includes work specialties as defined by duties and required skills as well as principal activity that a person does to earn money.

^c NeuroLex Educational assessment includes level of education and years of schooling.

^d NeuroLex Father's education assessment includes father's level of education and father's years of schooling.

^e NeuroLex Mother's education assessment includes mother's level of education and mother's years of schooling.

^f NeuroLex SAPS describes a type of assessment without any defined attributes.

^g NeuroLex SANS describes a type of assessment without any defined attributes.

data structure, it makes data available to consistent and efficient programmatic manipulation, reuse, transmission and storage, and it simplifies data conversion to other formats. XNAT comes with a set of XML schemas that describe common data associated with neuroimaging studies. XNAT also allows for the extension of these schemas as well as the creation of custom schemas. NUSDAST contains all three types: common, extended, and custom schemas. The extended and custom schemas include: subject registration data, extended demographic and relationship information, psychopathology measures based on SAPS and SANS, cognitive data and SNP data.

Custom subject registration schema

The registration information describes each subject's diagnostic category and longitudinal follow-up status and timing. We extended the standard XNAT base schema type subjectAssessorData to define a complex EncounterLog type schema. It is now a standard XNAT base schema available for any users of XNAT Central interested in preserving subject continuity information related to a study. The custom EncounterLog

schema describes data listed in **Supplemental Table ST1** and **Supplemental Figure SF1**.

Custom demographics and relationship schema

The standard XNAT base schema type demographicData was extended to define a complex nundaDemographicData type schema. The standard XNAT base demographic type describes Date of Birth, Year of Birth, Age, Gender, Handedness, Socioeconomic status, Education, Education description, Race, Ethnicity, Weight, Height, Gestational age, Post menstrual age, Birth weight. The extended nundaDemographicData and relationship schema describes additional data listed in **Supplemental Table ST2** and **Supplemental Figure SF2**.

Custom psychopathology schema

We extended the standard XNAT base schema type subjectAssessorData to define a complex symptomsSAPSSANS type schema. The custom symptomsSAPSSANS schema describes data listed in **Supplemental Table ST3** and **Supplemental Figure SF3**.

Custom cognition schema

We extended the standard XNAT base schema type subjectAssessorData to define a complex symptomsNeurocog type schema. The custom symptomsNeurocog schema describes data listed in **Supplemental Table ST4** and **Supplemental Figure SF4**.

Custom genetics schema

We created the genetic schema extension to accommodate schizophrenia-related genetic information. The custom genetics schema describes data listed in **Supplemental Table ST5** and **Supplemental Figure SF5**.

DATA STANDARDIZATION FOR SHARING WITHIN XNAT

In data sharing, it is essential for different projects with the same/similar data elements to use the same/similar schema, not different schemas, to describe the data. Toward this end, we first recognize that several of our assessments used industry-standard instruments that many other groups use, particularly in schizophrenia research. These instruments include the SAPS/SANS assessments for psychopathology and items of cognitive battery (e.g., WMS-III above). Custom schemas for these assessments have now been made standard within XNAT Central for ready use by other projects.

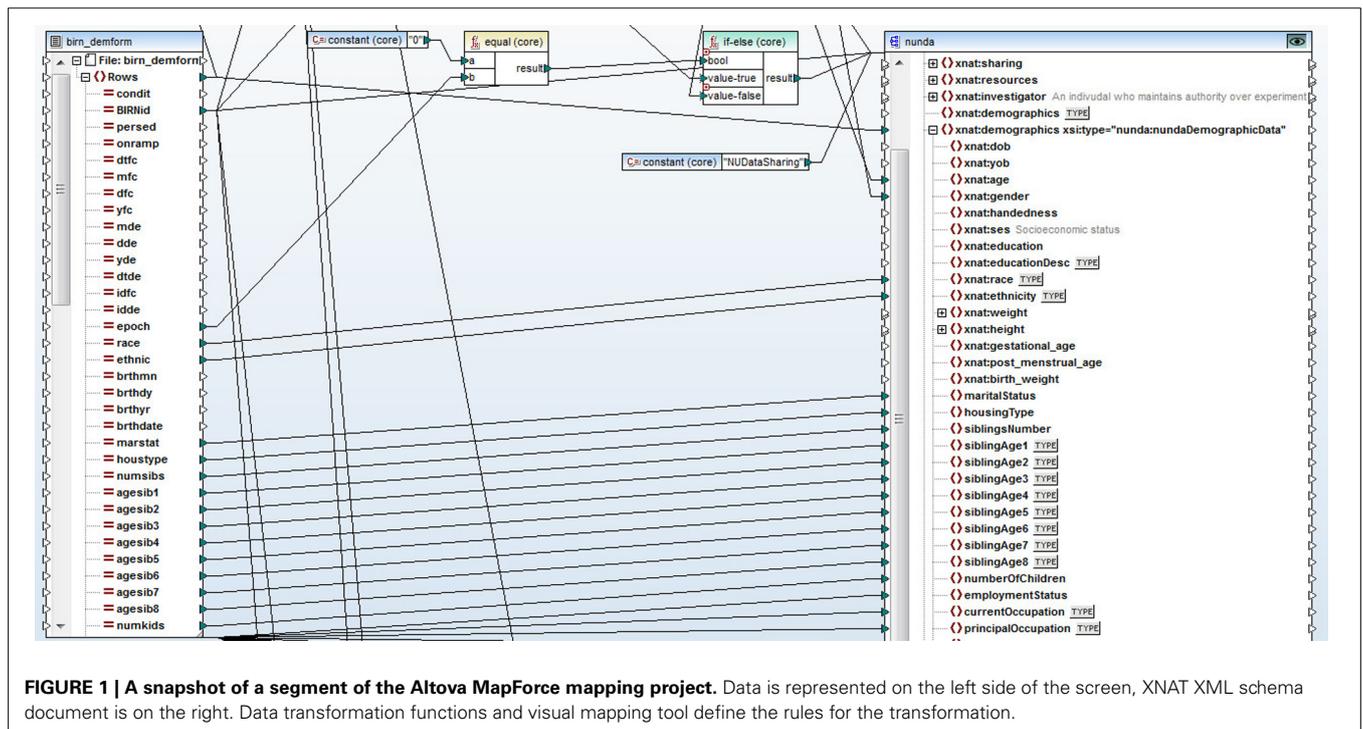
On the other hand, other data were obtained by non-standard assessment instruments but are nonetheless valuable for other projects. These data include the SNP (e.g., BDNF-rs6265 allele) and sibling relationships data (i.e., which two subjects are siblings and whether the sibling is a brother or sister). Custom schemas for these assessments are available in XNAT Central for ready adaptation by other projects.

Besides creating these new, custom schemas, we also took existing, XNAT standard schemas and extended to accommodate our data. These include the demographics and relationship schema. This extended schema is applicable to any XNAT Central project that preserves subject continuity information. We therefore have made it a standard base schema in XNAT Central.

DATA TRANSLATION AND UPLOAD INTO XNAT CENTRAL

The non-imaging data were stored in a number of different source architecture types such as excel spreadsheets, SAS and SPSS data files. We used Altova MapForce software (Altova GmbH, 2012) to convert the data into XML format. Altova MapForce provides the ability to create complex mappings between data in a variety of formats and the documents that define structure and relationships of data. In this case, we used Altova MapForce to transform the data into XML documents compliant with the XNAT (see XNAT Central below) XML schema extensions that we designed for each type of data. As a result, MapForce generated a set of XML files that contained demographic, clinical, symptomatic, registration and cognitive data. See **Figure 1** for a snapshot of the Altova MapForce mapping project segment, demonstrating data transformation functions and the visual mapping tool.

Our XNAT Central repository is entitled “NU Schizophrenia Data and Software Tool Federation using BIRN Infrastructure (NUSDAST).” Uploading the XML data files was facilitated by the use of Python library PyXNAT (<http://packages.python.org/pyxnat/>) (Schwartz, 2011; Schwartz et al., 2012), which is built on top of XNAT’s REST API. PyXNAT provides data management and the ability to connect with other existing neuroimaging software, such as FreeSurfer, to the data in the repository, as well as



the ability to distribute data to high performance computing clusters for performing CPU-intensive data analysis tasks. Non-XML data (such as surfaces and landmarks) were uploaded via cURL (<http://curl.haxx.se/>), also over the REST API. We are continuing to upload data as new schema extensions on the XNAT Central repository are being created and implemented.

DATA ACCESS

XNAT CENTRAL

The project “NU Schizophrenia Data and Software Tool Federation using BIRN Infrastructure (NUSDAST)” is hosted on XNAT Central (<http://tinyurl.com/av9h7jm>)² Within the NUSDAST project on the XNAT Central website, data are organized by subjects following the XNAT architecture: study registration, longitudinal epochs of MR sessions, symptoms, and cognitive battery (**Figure 2**). Within each epoch’s MR session, scans and associated segmentations, surfaces, landmarks, and other data are listed. User download is accomplished via the Download action or the Manage Files action displayed on the XNAT web page. Data can also be retrieved via the XNAT REST API.

²Full URL: https://central.xnat.org/app/template/XDATScreen_report_xnat_projectData.vm/search_element/xnat:projectData/search_field/xnat:projectData.ID/search_value/NUDataSharing

WEB-BASED DOMAIN-SPECIFIC PORTAL BUILT INTERFACING XNAT

We have created a web-based data portal (http://niacal.northwestern.edu/xnat_queries/NUSDAST) that acts as an interface between the user and the XNAT Central database. The data portal was designed to accommodate NUSDAST-specific queries as well as queries that will suit the user’s needs. While XNAT Central has an excellent search query interface, it is lacking certain important features that could facilitate domain-specific data query and access. One important point is that without registration at XNAT Central, one does not know whether NUSDAST contains their desired data. A very useful and important feature of the data portal is that anyone to obtain a summary count of their search criteria before registering at XNAT Central. For example, a search for “female” and “2 or more” MR sessions (i.e., longitudinal imaging data only) would return with “We have 73 subjects (164 MR imaging sessions) matching the parameters you specified.” The user then can decide to register with XNAT Central and obtain detailed data.

Another important feature of the portal is our ability to provide NUSDAST-specific information. For example, in an XNAT Central advanced search, the user would have to specify beforehand that he or she would like to have cognitive data even though it is not part of the search criteria, for otherwise cognitive data would not be included as part of the data download after the

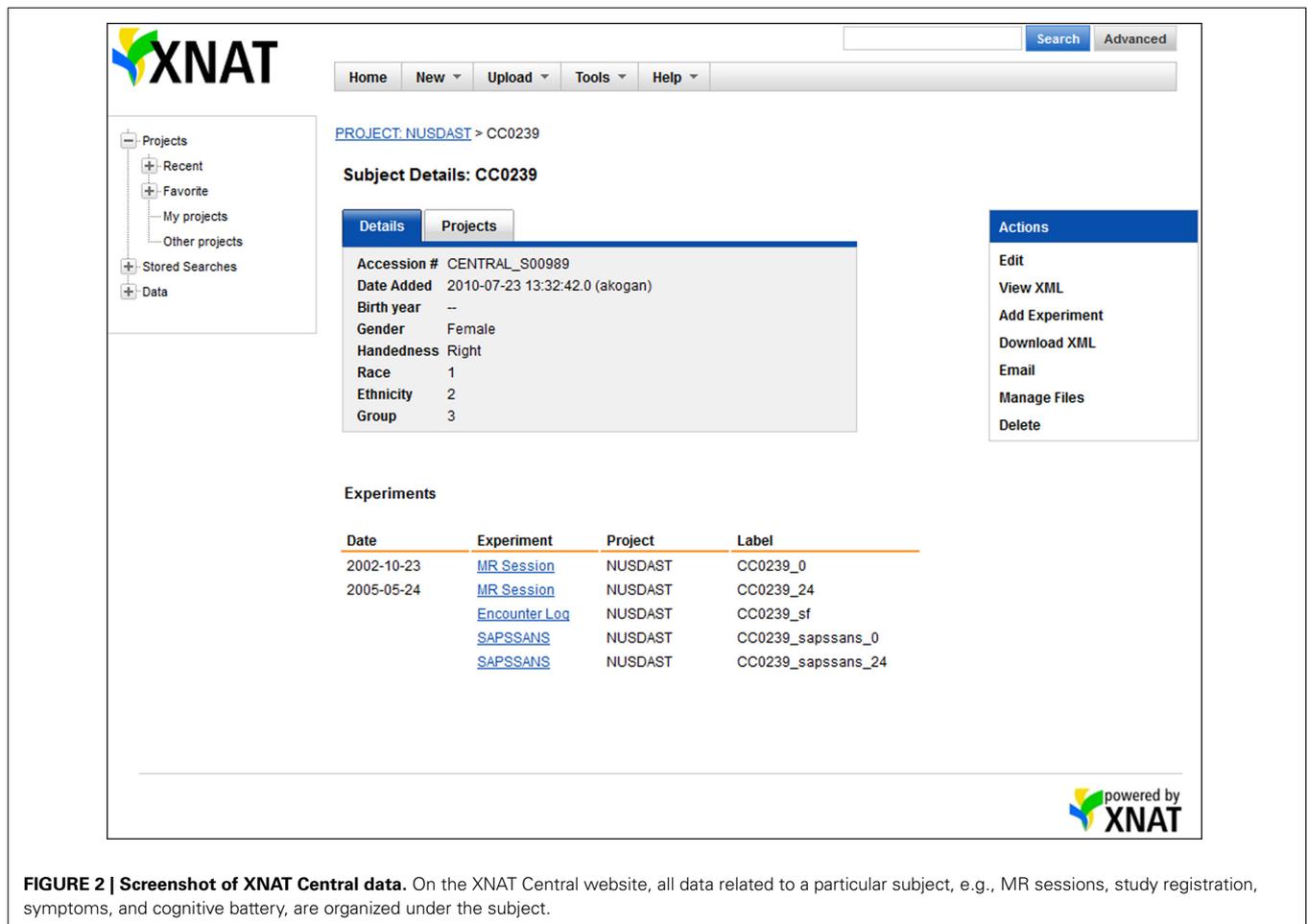


FIGURE 2 | Screenshot of XNAT Central data. On the XNAT Central website, all data related to a particular subject, e.g., MR sessions, study registration, symptoms, and cognitive battery, are organized under the subject.

search is complete. Our portal can be customized to give the user options of selecting what data they would like to have to download, after the subjects satisfying the search criteria are returned. For example, a user searching on age and imaging type parameters may want to download cognitive data for the subjects fitting search criteria.

The data portal was developed as part of a Ruby on Rails (<http://rubyonrails.org>) application and uses the HTTParty gem (<http://httparty.rubyforge.org/>) to communicate with the instance of XNAT via REST API. The portal presents options to specify search parameters for demographic, psychopathological, cognitive, genetic, and neuroimaging information that can be used to narrow the dataset which the user is looking to acquire. The application then sends a request to XNAT, in XML format, based on the search parameters specified and receives a list of data, also in XML format, that satisfies the specified criteria. The results are parsed, formatted and displayed on the data portal website along with instructions for downloading this data from XNAT Central. See **Figures 3, 4** for a dataflow diagram and a screenshot of the Data Portal. The user can also follow the XNAT Central link to browse and download NUSDAST data directly from XNAT Central. A download feature is available for data to be downloaded from the Data Portal (instead of directly interacting with XNAT Central), achieving a “one-click” capability.

NEUROIMAGING ANALYSIS TOOLS

CAWorks

The software application Computational Anatomy Works (CAWorks, **Figure 5**), available at <http://www.cis.jhu.edu/software/caworks/index.php>, was developed to support computational anatomy and shape analysis. CAWorks works seamlessly with the NUSDAST image, landmark and surface data and the Large Deformation Diffeomorphic Metric Mapping (LDDMM) mapping engine (Beg et al., 2005). The capabilities of CAWorks include:

- (1) Image and Shape Analysis plugin modules, such as Large Deformation Diffeomorphic Metric Mapping (LDDMM).
- (2) Interactive landmark placement to create segmentation (mask) of desired region of interest and specialized landmark placement plugins for the hippocampus, amygdala, and entorhinal cortex. After landmarking is completed, CAWorks facilitates submission for automated segmentation processing (e.g., LDDMM).
- (3) Browser plugin module for XNAT, enabling the retrieval of medical image data from XNAT for image and shape analysis and the storage of results in XNAT.
- (4) Quadra Planar view visualization.
- (5) Support for multiple Medical Imaging data formats, such as Nifti, Analyze, Freesurfer, DICOM, and landmark data.

CAWorks provides a GUI interface to any specified XNAT repository (including XNAT Central) such that the XNAT data appears as folders in a hierarchical structure (**Figure 5** panels A and B). When data is selected, it is cached locally and opened in the visualization window. The save option uploads and updates the XNAT repository. CAWorks gives researchers a more productive and effective experience in the analysis and visualization of data stored within XNAT. The direct interface with XNAT provides an intuitive hierarchical organizational structure which users may traverse and explore data. The user simply clicks on the data to be visualized and it appears in the interactive window. This saves the researcher the extra steps of uploading, downloading, and organizing the data temporarily on the file system.

DISCUSSION

In this paper, we described the development and deployment of a usable combination of schizophrenia-related dataset, tools for storing, sharing, support for computational anatomy and a data-mining portal, the NUSDAST.

Concerning data sharing, this resource built and extended upon existing, standard schemas available for data sharing on XNAT Central (<http://central.xnat.org/>). Specifically, we developed additional schemas for storing demographic, cognitive, genetic and clinical meta data in XNAT. These additions create

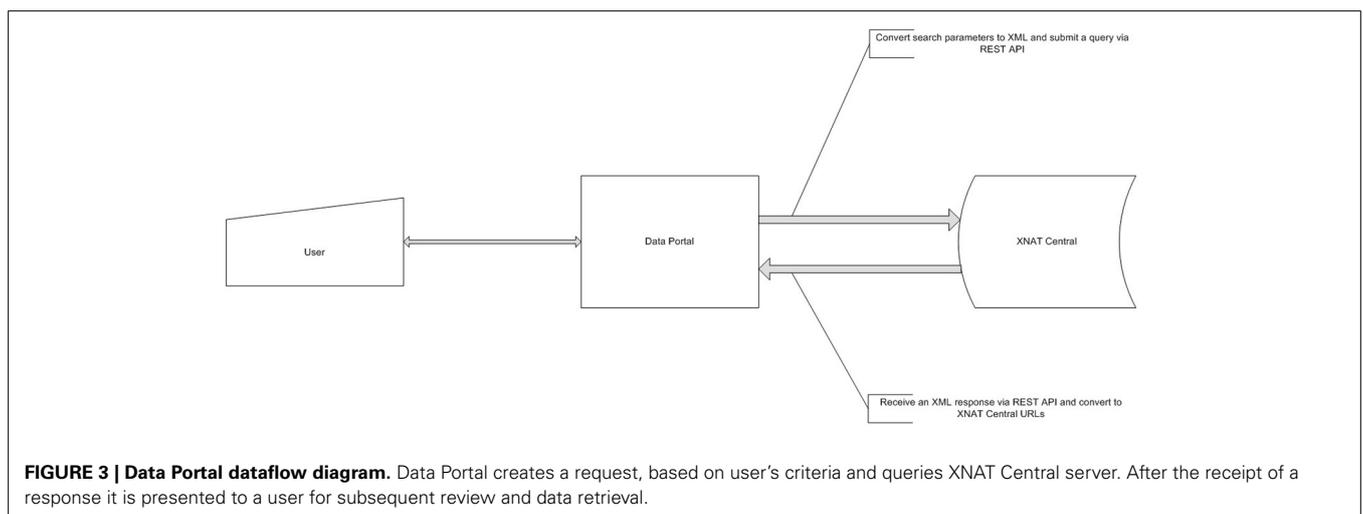


FIGURE 3 | Data Portal dataflow diagram. Data Portal creates a request, based on user's criteria and queries XNAT Central server. After the receipt of a response it is presented to a user for subsequent review and data retrieval.

FIGURE 4 | Screenshot of web data portal. Data Portal is used to search the collected dataset directly at the repository. Search results can be used to download data.

the opportunity to consistently expand and share schizophrenia research-related data³. We have significantly improved the way scientists are able to mine our dataset by creating a data portal for searching and downloading our data along with the accompanying longitudinal data.

Our well-described and comprehensive data on the normal controls and their siblings are valuable beyond the schizophrenia research community. For example, The SNP data include ones that are related to neurodevelopment (e.g., BDNF), embryonic development and tissue repair (e.g., FGF-2), and immune response/inflammation (e.g., EGFR, IL-3). Mutations of many of these SNPs have been found to be related to cancer and neuropsychiatric disorders such as depression, anxiety and Alzheimer

disease. Therefore, with the accompanying imaging and cognitive data, our control subjects data can be of wider utility beyond schizophrenia research.

Concerning data exploration, this source offers CAWorks, which extends ParaView, an open source, multi-platform, freely available program for parallel, interactive, scientific visualization. An important visualization tool for TeraGrid researchers, its client-server architecture facilitates remote visualization of datasets and the generation of level of detail (LOD) models that maintain interactive frame rates for large datasets. CAWorks is being used by consortium and center projects such as the Biomarkers for Older Controls At Risk for Dementia (BIOCARD), F.M. Kirby Research Center for Functional Brain Imaging, and BIRN projects. Additional functionality is being developed for storing and retrieving XNAT data as new analysis tools are added to CAWorks.

The schizophrenia research community has invested substantial resources in order to collect, manage and share increasingly larger datasets including neuroimaging data. The exploration of large, multi-modal, datasets has indeed improved our understanding of relationships among abnormalities of brain circuitry,

³Extended and new schemas created by NUSDAST can be used by other existing or future projects to describe their data. A benefit of sharing data through XNAT Central, rather than through an independent XNAT system, is that multiple projects can easily use the same schema representations and user interface components to represent common data elements. Users can then query and mine XNAT Central to locate data across the multiple projects using the shared schema.

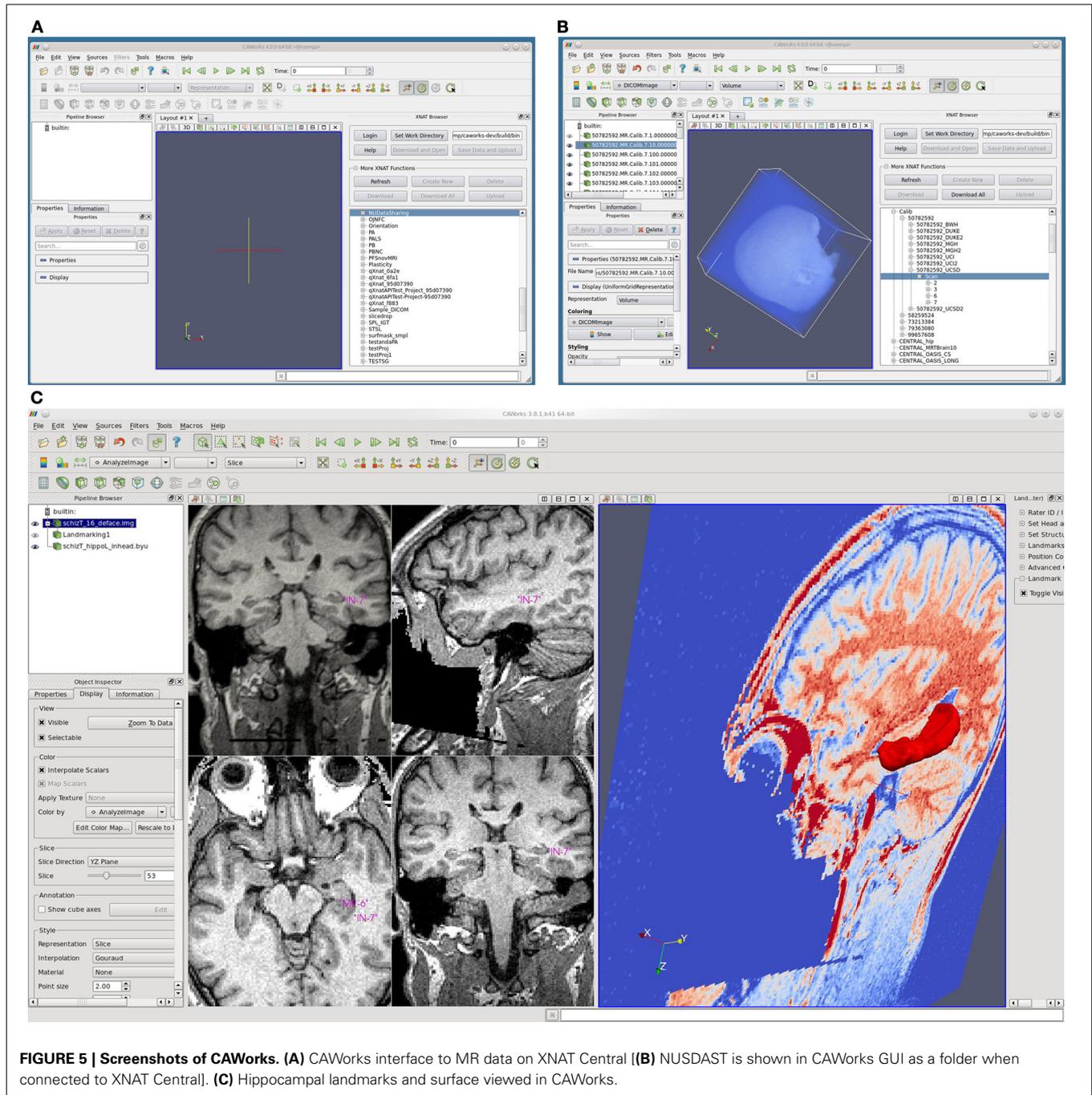


FIGURE 5 | Screenshots of CAWorks. (A) CAWorks interface to MR data on XNAT Central **(B)** NUSDAST is shown in CAWorks GUI as a folder when connected to XNAT Central. **(C)** Hippocampal landmarks and surface viewed in CAWorks.

brain function and genetic variability in schizophrenia (Kim et al., 2009, 2010; Allen et al., 2011).

Numerous data sharing initiatives were undertaken in order to create publicly accessible neuroimaging data collections, such as: FBIRN, MCIC, BSNIP, fMRI DataCenter (fMRIDC), the Open Access Series of Imaging Studies (OASIS). One of the main obstacles to the open-access sharing of research data is the lack of local organization and standard descriptions (Poline et al., 2012). Different resources usually organize data in many different formats, which leads to difficulties in sharing and analyzing data.

For example, the acquisition of a dataset from any given source will most likely require some programming work in order to make the dataset suitable for processing by an already utilized suite of tools. The concept of “one-click share” can only come true when appropriate standards are adopted and followed by the research community. The processes that will help establish standardized procedures include proliferation of open-source data sharing and of storage mechanisms. These systems have to respond to the current needs of researchers, to be comprehensive enough (or extensible) to accommodate different types of research

and to provide data-mining and common data processing functionality.

It is our hope that this effort will help to overcome some of the commonly recognized technical barriers in advancing neuroimaging research (Poline et al., 2012) by creating a research-ready dataset that meaningfully combines neuroimaging data with other relevant information. Currently, more data are being made available, such as fMRI and genome-wide scan (GWS) data. We are also beginning to expand the scope of schizophrenia neuroimaging data sharing by linking NUSDAST with FBIRN and MCIC through the development of SchizConnect, which will be a data mediation and integration platform that establishes a true federation of disparate, heterogeneous neuroimaging-related databases.

ACKNOWLEDGMENTS

This work was supported in part by NIH grants 1R01 MH084803, 1U01 MH097435, P50 MH071616, R01 MH056584, P41 RR15241, and by the NIH through the following NIGMS grant: U24 GM104203 Bio-Informatics Research Network Coordinating Center (BIRN-CC).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://www.frontiersin.org/journal/10.3389/fninf.2013.00025/abstract>

Supplemental Figure SF1 | Custom EncounterLog schema.

Supplemental Figure SF2 | Extended nundaDemographicData and relationship schema.

Supplemental Figure SF3 | Custom symptomsSAPSSANS schema.

Supplemental Figure SF4 | Custom symptomsNeurocog schema.

Supplemental Figure SF5 | Custom genetics schema.

Supplemental Table ST1 | Custom EncounterLog schema.

Supplemental Table ST2 | Extended nundaDemographicData and relationship schema.

Supplemental Table ST3 | Custom symptomsSAPSSANS schema.

Supplemental Table ST4 | Custom symptomsNeurocog schema.

Supplemental Table ST5 | Custom genetics schema.

REFERENCES

- Allen, E. A., Erhardt, E. B., Damaraju, E., Gruner, W., Segall, J. M., Silva, R. F., et al. (2011). A baseline for the multivariate comparison of resting-state networks. *Front. Syst. Neurosci.* 5:2. doi: 10.3389/fnsys.2011.00002
- Altova GmbH. (2012). Altova MapForce.
- Andreasen, N. C. (1983). *The Scale for Assessment of Negative Symptoms (SANS)*. Iowa City: The University of Iowa.
- Andreasen, N. C. (1984). *The Scale for Assessment of Positive Symptoms (SAPS)*. Iowa City: The University of Iowa.
- Andreasen, N. C., Arndt, S., Alliger, R., Miller, D., and Flaum, M. (1995). Symptoms of schizophrenia. Methods, meanings, and mechanisms. *Arch. Gen. Psychiatry* 52, 341–351. doi: 10.1001/archpsyc.1995.03950170015003
- Barch, D. M., Csernansky, J., Conturo, T., Snyder, A. Z., and Ollinger, J. (2002). Working and long-term memory deficits in schizophrenia. Is there a common underlying prefrontal mechanism? *J. Abnorm. Psychol.* 111, 478–494. doi: 10.1037/0021-843X.111.3.478
- Beg, M. F., Miller, M. I., Troune, A., and Younes, L. (2005). Computing large deformation metric mappings via geodesic flows of diffeomorphisms. *Int. J. Comput. Vis.* 61, 139. doi: 10.1023/B:VISI.0000043755.93987.aa
- Benton, A. L. (1976). *Multilingual Aphasia Examination*. Iowa: University of Iowa.
- Benton, A. L., Hamsher, K. D. S., and Sivan, A. B. (1984). *Multilingual Aphasia Examination (MAE)*, 3rd Edn. Iowa: Psychological Assessment Resources.
- Bockholt, H. J., Scully, M., Courtney, W., Rachakonda, S., Scott, A., Caprihan, A., et al. (2010). Mining the mind research network: a novel framework for exploring large scale, heterogeneous translational neuroscience research data sources. *Front. Neuroinform.* 3:36. doi: 10.3389/neuro.11.036.2009
- Braver, T. S., Cohen, J. D., Nystrom, L. E., Jonides, J., Smith, E. E., and Noll, D. C. (1997). A parametric study of prefrontal cortex involvement in human working memory. *Neuroimage* 5, 49–62. doi: 10.1006/nimg.1996.0247
- Buckner, R. L., Head, D., Parker, J., Fotenos, A. F., Marcus, D., Morris, J. C., et al. (2004). A unified approach for morphometric and functional data analysis in young, old, and demented adults using automated atlas-based head size normalization: reliability and validation against manual measurement of total intracranial volume. *Neuroimage* 23, 724–738. doi: 10.1016/j.neuroimage.2004.06.018
- Claudio, M., and Roberto, S. (1994). Using Marching Cubes on small machines. *Graph. Models Image Process.* 56, 182–183. doi: 10.1006/cgip.1994.1017
- Cobia, D. J., Csernansky, J. G., and Wang, L. (2011). Cortical thickness in neuropsychologically near-normal schizophrenia. *Schizophr. Res.* 133, 68–76. doi: 10.1016/j.schres.2011.08.017
- Cornblatt, B. A., Risch, N. J., Faris, G., Friedman, D., and Erlenmeyer-Kimling, L. (1988). The Continuous Performance Test, identical pairs version (CPT-IP): I. New findings about sustained attention in normal families. *Psychiatry Res.* 26, 223–238. doi: 10.1016/0165-1781(88)90076-5
- Csernansky, J. G., and Bardgett, M. E. (1998). Limbic-cortical neuronal damage and the pathophysiology of schizophrenia. *Schizophr. Bull.* 24, 231–248. doi: 10.1093/oxfordjournals.schbul.a033323
- Csernansky, J. G., Schindler, M. K., Splinter, N. R., Wang, L., Gado, M., Selemon, L. D., et al. (2004a). Abnormalities of thalamic volume and shape in schizophrenia. *Am. J. Psychiatry* 161, 896–902. doi: 10.1176/appi.ajp.161.5.896
- Csernansky, J. G., Wang, L., Joshi, S. C., Ratnanather, J. T., and Miller, M. I. (2004b). Computational anatomy and neuropsychiatric disease: probabilistic assessment of variation and statistical inference of group difference, hemispheric asymmetry, and time-dependent change. *Neuroimage* 23 (Suppl 1), S56–68. doi: 10.1016/j.neuroimage.2004.07.025
- Csernansky, J. G., Wang, L., Jones, D., Rastogi-Cruz, D., Posener, J. A., Heydebrand, G., et al. (2002). Hippocampal deformities in schizophrenia characterized by high dimensional brain mapping. *Am. J. Psychiatry* 159, 2000–2006. doi: 10.1176/appi.ajp.159.12.2000
- Delawalla, Z., Barch, D. M., Fisher Eastep, J. L., Thomason, E. S., Hanewinkel, M. J., Thompson, P. A., et al. (2006). Factors mediating cognitive deficits and psychopathology among siblings of individuals with schizophrenia. *Schizophr. Bull.* 32, 525–537. doi: 10.1093/schbul/sbj082
- Desikan, R. S., Segonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., et al. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* 31, 968–980. doi: 10.1016/j.neuroimage.2006.01.021
- Duvernoy, H. M. (1988). *The Human Hippocampus: An Atlas of Applied Anatomy*. Munich: J. F. Bergmann Verlag. doi: 10.1007/978-3-642-54195-7
- Duvernoy, H. M. (1991). *The Human Brain: Surface, Three-Dimensional Anatomy and MRI*. Wien, NY: Springer-Verlag.
- Friedman, L., Stern, H., Brown, G. G., Mathalon, D. H., Turner, J., Glover, G. H., et al. (2008). Test-retest and between-site reliability in a multicenter fMRI study. *Hum. Brain Mapp.* 29, 958–972. doi: 10.1002/hbm.20440
- Goldman-Rakic, P. S. (1999). The physiological approach: functional architecture of working memory and disordered cognition in schizophrenia. *Biol. Psychiatry* 46, 650–661. doi: 10.1016/S0006-3223(99)00130-4
- Grenander, U., and Miller, M. I. (1998). Computational anatomy: an emerging discipline. *Q. Appl. Math.* LVI, 617–694.

- Gur, R. C., Ragland, J. D., Moberg, P. J., Bilker, W. B., Kohler, C., Siegel, S. J., et al. (2001). Computerized neurocognitive scanning: II. The profile of schizophrenia. *Neuropsychopharmacology* 25, 777–788. doi: 10.1016/S0893-133X(01)00279-2
- Haller, J. W., Banerjee, A., Christensen, G. E., Gado, M., Joshi, S., Miller, M. I., et al. (1997). Three-dimensional hippocampal MR morphometry with high-dimensional transformation of a neuroanatomic atlas. *Radiology* 202, 504–510.
- Harms, M. P., Wang, L., Campanella, C., Aldridge, K., Moffitt, A. J., Kuelper, J., et al. (2010). Structural abnormalities in gyri of the prefrontal cortex in individuals with schizophrenia and their unaffected siblings. *Br. J. Psychiatry* 196, 150–157. doi: 10.1192/bjp.bp.109.067314
- Harms, M. P., Wang, L., Mamah, D., Barch, D. M., Thompson, P. A., and Csernansky, J. G. (2007). Thalamic shape abnormalities in individuals with schizophrenia and their nonpsychotic siblings. *J. Neurosci.* 27, 13835–13842. doi: 10.1523/JNEUROSCI.2571-07.2007
- Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., and Curtiss, G. (1993). *Wisconsin Card Sort-ing Test Manual*. Odessa: Psychological Assessment Resources, Inc.
- Helmer, K. G., Ambite, J. L., Ames, J., Ananthakrishnan, R., Burns, G., Chervenak, A. L., (2011). Enabling collaborative research using the Biomedical Informatics Research Network (BIRN). *J. Am. Med. Inform. Assoc.* 18, 416–422. doi: 10.1136/amiajnl-2010-000032
- Jacoby, L. L. (1999). Deceiving the elderly: effects of accessibility bias in cued-recall performance. *Cogn. Neuropsychol.* 16, 417–436. doi: 10.1080/026432999380861
- Jacoby, L. L., Debner, J. A., and Hay, J. F. (2001). Proactive interference, accessibility bias, and process dissociations: valid subjective reports of memory. *J. Exp. Psychol. Learn. Mem. Cogn.* 27, 686–700. doi: 10.1037/0278-7393.27.3.686
- Jacoby, L. L., Ste-Marie, D., and Toth, J. P. (1993a). “Redefining automaticity: unconscious influences, awareness, and control,” in *Attention: Selection, Awareness, and Control: A Tribute to Donald Broadbent*, eds A. Baddeley and L. Weiskrantz. (New York, NY: Clarendon), 261–282.
- Jacoby, L. L., Toth, J. P., and Yonelinas, A. P. (1993b). Separating conscious and unconscious influences of memory: measuring recollection. *J. Exp. Psychol. Gen.* 122, 139–154. doi: 10.1037/0096-3445.122.2.139
- John, J. P., Wang, L., Moffitt, A. J., Singh, H. K., Gado, M. H., and Csernansky, J. G. (2006). Inter-rater reliability of manual segmentation of the superior, inferior and middle frontal gyri. *Psychiatry Res.* 148, 151–163. doi: 10.1016/j.psychres.2006.05.006
- Joshi, S., Miller, M. I., Christensen, G. E., Coogan, T. A., and Grenander, U. (1995). The generalized dirichlet problem for mapping brain manifolds. *Int. Symp. Opt. Sci. Eng. Instr.* 278–289.
- Joshi, S., Miller, M. I., and Grenander, U. (1997). On the geometry and shape of brain sub-manifolds. *Int. J. Pattern. Recognit. Artif. Intell.* 11, 1317–1343. doi: 10.1142/S0218001497000615
- Karnik-Henry, M. S., Wang, L., Barch, D. M., Harms, M. P., Campanella, C., and Csernansky, J. G. (2012). Medial temporal lobe structure and cognition in individuals with schizophrenia and in their non-psychotic siblings. *Schizophr. Res.* 138, 128–135. doi: 10.1016/j.schres.2012.03.015
- Keator, D. B., Grethe, J. S., Marcus, D., Ozyurt, B., Gadde, S., Murphy, S., et al. (2008). A national human neuroimaging collaborative enabled by the Biomedical Informatics Research Network (BIRN). *IEEE Trans. Inf. Technol. Biomed.* 12, 162–172. doi: 10.1109/TITB.2008.917893
- Kim, D. I., Manoach, D. S., Mathalon, D. H., Turner, J. A., Mannell, M., Brown, G. G., et al. (2009). Dysregulation of working memory and default-mode networks in schizophrenia using independent component analysis, an fBIRN and MCIC study. *Hum. Brain Mapp.* 30, 3795–3811. doi: 10.1002/hbm.20807
- Kim, M. A., Tura, E., Potkin, S. G., Fallon, J. H., Manoach, D. S., Calhoun, V. D., et al. (2010). Working memory circuitry in schizophrenia shows widespread cortical inefficiency and compensation. *Schizophr. Res.* 117, 42–51. doi: 10.1016/j.schres.2009.12.014
- Lorensen, W. E., and Cline, H. E. (1987). Marching cubes: a high resolution 3d surface construction algorithm. *Comput. Graph.* 21, 163–169. doi: 10.1145/37402.37422
- Mai, J. K., Assheuer, J., and Paxinos, G. (1997). *Atlas of the Human Brain*. San Diego: Academic Press.
- Mamah, D., Wang, L., Barch, D., de Erausquin, G. A., Gado, M., and Csernansky, J. G. (2007). Structural analysis of the basal ganglia in schizophrenia. *Schizophr. Res.* 89, 59–71. doi: 10.1016/j.schres.2006.08.031
- Marcus, D. S., Olsen, T. R., Ramaratnam, M., and Buckner, R. L. (2007a). The Extensible Neuroimaging Archive Toolkit: an informatics platform for managing, exploring, and sharing neuroimaging data. *Neuroinformatics* 5, 11–34. doi: 10.1385/NI:5:1:11
- Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., and Buckner, R. L. (2007b). Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. *J. Cogn. Neurosci.* 19, 1498–1507. doi: 10.1162/jocn.2007.19.9.1498
- Nuechterlein, K. H., Barch, D. M., Gold, J. M., Goldberg, T. E., Green, M. F., and Heaton, R. K. (2004). Identification of separable cognitive factors in schizophrenia. *Schizophr. Res.* 72, 29–39. doi: 10.1016/j.schres.2004.09.007
- Poline, J. B., Breeze, J. L., Ghosh, S., Gorgolewski, K., Halchenko, Y. O., Hanke, M., et al. (2012). Data sharing in neuroimaging research. *Front. Neuroinform.* 6:9. doi: 10.3389/fninf.2012.00009
- Potkin, S. G., and Ford, J. M. (2009). Widespread cortical dysfunction in schizophrenia: the FBIRN imaging consortium. *Schizophr. Bull.* 35, 15–18. doi: 10.1093/schbul/sbn159
- Potkin, S. G., Macciardi, F., Guffanti, G., Fallon, J. H., Wang, Q., Turner, J. A., et al. (2010). Identifying gene regulatory networks in schizophrenia. *Neuroimage* 53, 839–847. doi: 10.1016/j.neuroimage.2010.06.036
- Potkin, S. G., Turner, J. A., Brown, G. G., McCarthy, G., Greve, D. N., Glover, G. H., et al. (2009). Working memory and DLPFC inefficiency in schizophrenia: the FBIRN study. *Schizophr. Bull.* 35, 19–31. doi: 10.1093/schbul/sbn162
- Qiu, A., Younes, L., Wang, L., Ratnanather, J. T., Gillepsie, S. K., Kaplan, G., et al. (2007). Combining anatomical manifold information via diffeomorphic metric mappings for studying cortical thinning of the cingulate gyrus in schizophrenia. *Neuroimage* 37, 821–833. doi: 10.1016/j.neuroimage.2007.05.007
- Rastogi-Cruz, D., and Csernansky, J. (1997). “Clinical rating scales,” in *Adult Psychiatry*, ed S. Guze (St. Louis, MO: Mosby, Inc.), 45–51.
- Reitan, R. M., and Wolfson, D. (1985). *The Halstead-Reitan Neuropsychological Test Battery: Theory and Clinical Interpretation*. Tucson, AZ: Neuropsychology Press.
- Ripke, S., Sanders, A. R., Kendler, K. S., Levinson, D. F., Sklar, P., Holmans, P. A., et al. (2011). Genome-wide association study identifies five new schizophrenia loci. *Nat. Genet.* 43, 969–976. doi: 10.1038/ng.940
- Schwartz, Y. (2011). PyXNAT.
- Schwartz, Y., Barbot, A., Thyreau, B., Frouin, V., Varoquaux, G., Siram, A., et al. (2012). PyXNAT: XNAT in Python. *Front. Neuroinform.* 6:12. doi: 10.3389/fninf.2012.00012
- Smith, M. J., Wang, L., Cronenwett, W., Mamah, D., Barch, D. M., and Csernansky, J. G. (2011). Thalamic morphology in schizophrenia and schizoaffective disorder. *J. Psychiatr. Res.* 45, 378–385. doi: 10.1016/j.jpsychires.2010.08.003
- Tamminga, C. A., Ivleva, E. I., Keshavan, M. S., Pearson, G. D., Clementz, B. A., Witte, B., et al. (2013). Clinical phenotypes of psychosis in the bipolar and schizophrenia network on intermediate phenotypes (B-SNIP). *Am. J. Psychiatry* doi: 10.1176/appi.ajp.2013.12101339. [Epub ahead of print].
- Turner, J. A., Calhoun, V. D., Michael, A., van Erp, T. G. M., Ehrlich, S., Segall, J. M., et al. (2012). Heritability of multivariate gray matter measures in schizophrenia. *Twin. Res. Hum. Genet.* 15, 324–335. doi: 10.1017/thg.2012.1
- Venkatesan, R., and Haacke, E. (1997). Role of high resolution in magnetic resonance (MR) imaging: applications for MR angiography, intracranial T1-weighted imaging, and image interpolation. *Int. J. Imaging Syst. Technol.* 8, 529–543. doi: 10.1002/(SICI)1098-1098(1997)8:6<529::AID-IMA5>3.3.CO;2-M
- Wang, L., Hosakere, M., Trein, J. C., Miller, A., Ratnanather, J. T., Barch, D. M., et al. (2007a). Abnormalities of cingulate gyrus neuroanatomy in schizophrenia. *Schizophr. Res.* 93, 66–78. doi: 10.1016/j.schres.2007.02.021
- Wang, L., Lee, D. Y., Bailey, E., Hartlein, J. M., Gado, M. H., Miller, M. I., et al. (2007b). Validity of large-deformation high dimensional brain mapping of the basal ganglia in adults with Tourette syndrome. *Psychiatry Res.* 154, 181–190. doi: 10.1016/j.psychres.2006.08.006
- Wang, L., Joshi, S. C., Miller, M. I., and Csernansky, J. G. (2001). Statistical analysis of hippocampal asymmetry in schizophrenia. *Neuroimage* 14, 531–545. doi: 10.1006/nimg.2001.0830
- Wang, L., Mamah, D., Harms, M. P., Karnik, M., Price, J. L., Gado, M. H., et al. (2008). Progressive deformation of deep brain nuclei and

hippocampal-amygdala formation in schizophrenia. *Biol. Psychiatry* 64, 1060–1068. doi: 10.1016/j.biopsych.2008.08.007

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale*. San Antonio, TX: The Psychological Corporation.

Weinberger, D. R., Berman, K. F., Suddath, R., and Torrey, E. F. (1992). Evidence of dysfunction of a prefrontal-limbic network in schizophrenia: a magnetic resonance imaging and regional cerebral blood flow study of discordant monozygotic twins. *Am. J. Psychiatry* 149, 890–897.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 08 May 2013; paper pending published: 17 June 2013; accepted: 12 October 2013; published online: 07 November 2013.

Citation: Wang L, Kogan A, Cobia D, Alpert K, Kolasny A, Miller MI and Marcus D (2013) Northwestern University Schizophrenia Data and Software Tool (NUSDAST). *Front. Neuroinform.* 7:25. doi: 10.3389/fninf.2013.00025

This article was submitted to the journal *Frontiers in Neuroinformatics*.

Copyright © 2013 Wang, Kogan, Cobia, Alpert, Kolasny, Miller and Marcus. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.