

83 Neuroimaging Informatics: Tools to Manage and Share Neuroimaging and Related Data

DANIEL MARCUS

ABSTRACT Neuroimaging informatics is the set of data-centric technologies used to enable the practice of neuroimaging-based science. It includes technologies like data management platforms, databases of stored data and knowledge, data structures for representing images and other data, and a variety of software applications to achieve everything from data capture to high-throughput analysis to data mining. In this chapter, I review the current state of the art in neuroimaging informatics, frequently using the XNAT informatics platform and Human Connectome Project as examples.

Why neuroimaging informatics?

The Human Connectome Project (HCP), a National Institutes of Health-funded project to generate connectivity maps of the human brain using state-of-the-art neuroimaging methods, will study 1,200 individuals over a five-year period, including collecting structural, functional, and diffusion magnetic resonance imaging (MRI), extensive behavioral and cognitive phenotypic data, and next-generation genomics. While the HCP is currently at the cutting edge of neuroimaging capabilities, its methods are quickly being adopted and incorporated into a broad range of other studies. It therefore serves as a useful example of the sort of data challenges that large-scale neuroimaging studies encounter:

1. *Big data.* For each HCP subject, the acquired imaging data is over 10GB, the preprocessed data is over 12GB, and the dense connectome data is over 50GB (Marcus et al., 2013). In total, the HCP data set is expected to exceed 1,000TB, or the equivalent of over 220,000 DVDs.

2. *Multimodal data.* The HCP imaging protocol includes T1- and T2-weighted structural scans, multiple resting-state functional MRI (fMRI) scans, eight separate task fMRI scans, and diffusion imaging. The scans are acquired in four to five separate imaging sessions (Van Essen et al., 2013). During preprocessing, the

modalities are spatially coregistered, and many analytic approaches will synthesize across modalities.

3. *Nonimaging measures.* The HCP behavioral battery includes dozens of assessments and tests, covering cognitive, emotional, and sensory domains (Barch et al., 2013). The battery is acquired on several different computerized testing platforms, each using its own proprietary data format.

4. *Extensive image processing.* The acquired imaging data is processed following a standardized sequence of steps that includes coregistration, distortion correction, denoising, and surface reconstruction (Glasser et al., 2013). On the HCP's compute cluster, each subject's preprocessing requires over 24 hours to execute.

5. *Sensitive subject information.* The high-resolution images acquired by the HCP contain facial characteristics of study participants that may be considered personal health information (PHI) (Chen et al., 2007). In addition, many of the behavioral measures may be considered highly personal by subjects and their families, including drug use, pregnancy status, and psychiatric traits. Human subject protections and federal Health Insurance Portability and Accountability Act (HIPAA) regulations require subject privacy and confidentiality to be maintained when distributing these data.

6. *Collaboration.* The HCP research team consists of 100 personnel from 100 institutions in five countries. In order for this collaborative team to work effectively, the data must be made securely accessible across a geographically dispersed network.

7. *Data sharing.* The HCP grant carries a mandate from the NIH to proactively share its data as openly as possible with the international neuroimaging community.

Given these challenges, not surprisingly, a number of informatics systems have emerged to provide data management, security, sharing, and workflow solutions (Das, Zijdenbos, Harlap, Vins, & Evans, 2011; Ozyurt

et al., 2010; Scott et al., 2011). The XNAT imaging informatics platform, developed in my own laboratory, was designed to address many of these issues (Marcus, Olsen, Ramaratnam, & Buckner, 2007). I will use XNAT throughout this chapter to illustrate the sort of informatics capabilities that are important for managing large-scale neuroimaging studies. XNAT is open source and extensible and is widely used in the neuroimaging community, including as the HCP's internal private database and open-access public database.

The XNAT platform is built on a standardized workflow that is intended to mirror the real-world operations of collaborative research studies such as the HCP. The workflow covers a stepwise process from acquisition through data sharing and interweaves the best practices, policy, and sociology of doing neuroimaging science. A quarantine stage, for example, provides a mechanism for reviewing data quality and completeness prior to use by local investigators. Additional stages enable secure data access by specific collaborators and by the broader research community. With each stage, a variety of productivity tools, such as web-based reports, searching tools, and automated processing routines, are provided to facilitate use of the data.

Data organization

Most neuroimaging informatics systems, including XNAT, model the data using a hierarchical organization. The *project level* contains data that are related to one another. Often a project is equivalent to a research study, but it could also be used to hold a subset of data from a study or a superset of data aggregated from multiple studies. Within the HCP, for example, a number of different projects are used to manage various subcomponents of the study, including 3 Tesla (3T) optimization data, 7 Tesla (7T) optimization data, and the primary study. Regardless of the contents of a project, the primary role of the project in the data hierarchy is to provide an element around which data can be conveniently grouped for navigation and for delimiting user access privileges. The *subject level* of the hierarchy represents the individuals on whom measurements are made, usually a human but also nonhuman primates and other model species. The subject level includes demographic and other nonchanging information about the subject and contains the set of experimental data elements obtained from the individual. The *visit level* captures all data obtained from a subject across all methods (e.g., MRI session, positron emission tomography session, neuropsychological evaluation) within a scoped time frame. Visits are often completed in a single day or two but may be open for weeks or

months in complex studies. Within the HCP, all subjects undergo a primary visit on the 3T scanner; a subset of subjects will undergo a second 7T visit at a later date. The visit level is especially useful for tracking of data within longitudinal studies that obtain repeated measures over time. The *experiment level* contains the actual experimental data acquired in the study, with specific experiment types defined to capture the specific elements associated with a particular instrument. For example, XNAT includes data experiment types for all Digital Imaging and Communications in Medicine (DICOM) imaging modalities (PET, MRI, CT, etc.) and many common psychological instruments. Custom experiment types can be added to XNAT systems to capture additional instruments.

Much of the functionality of XNAT and similar systems is built around this type of four-level data hierarchy; it enables experimental data to be navigated, grouped, tracked, and validated. For example, XNAT includes a protocol-tracking feature that allows study managers to define the expected timing of visits and the types of data to be collected within each visit. From this definition, XNAT provides user interfaces for entering the data associated with a visit and for navigating existing subject data.

User and programming interfaces

Neuroimaging informatics systems typically provide a web-based user interface. XNAT's web interface enables users to navigate by data type and by project. When navigating by data type, all data of that type (that the user is authorized to view) from across projects are aggregated into a single view. When navigating by project, data of different types within a single project are aggregated together. In the project-based view, users can review the various types of data within the project and administer project settings. From either navigation path, users can select specific subjects and experiments to view in more detail. At each level of navigation, users are presented with a set of context-specific actions for further interacting with the data. From the aggregated data tables, for example, users can download spreadsheets and filter the presented data. From individual experiment reports, they can generate PDF views and email links to the report to colleagues. If a user has sufficient access privileges, they are also presented with options to execute processing pipelines and edit the data.

In addition to the web-based user interface, systems often provide application programming interfaces (APIs) for programmatically interacting with the database. The XNAT API follows a representational state

transfer (REST) web-services architecture, which provides distinct and knowable hypertext transfer protocol (HTTP)-based resource locators for all hosted data elements and files (Fielding & Taylor, 2002). The data-access portion of the API closely follows XNAT's data hierarchy, including providing access to custom data types added to an XNAT repository using .xsd file extensions. The API also includes administration, pipeline, and operations components to provide comprehensive interactivity with XNAT functionality. A primary usage of the API is to enable external software to interface with XNAT. The HCP visualization system, Connectome Workbench, for example, uses the XNAT REST API to retrieve images for visualization and to execute dynamic multisubject analyses. Similarly, the API can be used to write scripts that interface with XNAT; scripting libraries have been developed in a number of languages, including Java, R, MATLAB, and Python.

Importing images

The core of a neuroimaging study, of course, is the data acquired at the scanner. In recent years, capturing these data directly from the scanner has been greatly eased by the near-universal implementation of the DICOM standard developed by the scanner manufacturers (National Electrical Manufacturers Association, 2011). The DICOM standard defines both a data format and a network transmission protocol. The data format includes the actual image data as well as a rich set of metadata that details how the images were obtained. Reviewing these metadata is an important component of a rigorous quality control process (see below), and a handful of the fields are essential for subsequent processing and analysis of the images. The DICOM network protocol specifies how image data is sent from the scanner (and other DICOM-based systems) to DICOM-compatible receivers. The most common receivers are the picture archiving and communication systems designed primarily for clinical use. Because they are designed for clinical use, picture archiving and communication systems lack a number of features—longitudinal views, integration, post-processing, data access control, and so on—that are important for organizing and managing neuroimaging research data. A main focus of neuroimaging informatics systems is to implement these functions. XNAT, for example, includes a DICOM receiver and metadata import system that maps incoming DICOM data to study-specific projects, around which security and data access privileges are constructed, and to research subjects and longitudinal study visits. XNAT also provides a mapping to bridge across the varying terminologies used in the

clinically focused DICOM standard and by research neuroimagers. In typical practice, a DICOM “study” is equivalent to a research “session,” and a “series” is equivalent to a “scan.”

In research neuroimaging, .nifti has emerged as the de facto standard file format. However, no scanners produce it, and so it is necessarily a derivative product generated through a file conversion process. The NIfTI metadata model is significantly more limited than DICOM, and much useful data acquisition information is not preserved in the conversion process (e.g., repetition time, flip angle, etc.). As this information is lost, many researchers have fallen into the trap of *assuming* they know the acquisition parameters for their data. Inevitably, with this approach, acquisition protocols will drift over time as scanners are upgraded or they are “tweaked” by colleagues. To avoid this pitfall, it's important to preserve the original DICOM files and to implement quality control and review procedures (see below). Despite this limitation, NIfTI has several important attributes. Most importantly, it implements a coherent coordinate system that accounts for left-right directionality in the data and supports up to 128-bit floating-point binary data. It also includes several convenience features like compact single file storage and an extension mechanism to incorporate additional data or metadata.

While DICOM is the preferred format for archiving data (given the pitfalls described above), neuroinformatics platforms often implement a NIfTI-based data import workflow in addition to or instead of DICOM. Because NIfTI does not define a network-transmission protocol, nor does it include metadata by which study organization can be inferred, data systems must implement data-transfer procedures and mechanisms for inferring data organization. XNAT, for example, includes a NIfTI web-based upload interface that imports the data into projects, sessions, or scans according to directory and file names. Users can customize the naming patterns to match their data. Acquisition metadata for the imported NIfTI data can be supplied using the XNAT programming interface.

Quality control

As neuroimaging studies continue to grow in scale and complexity, the risk of data-collection errors grows as well. Such errors include systematic acquisition with an incorrect protocol or sporadic errors like poor head positioning, subject motion, and susceptibility artifacts. Quality control (QC) procedures built into the informatics workflow can be used to mitigate these risks. QC procedures include image-acquisition validation,

manual image review, and automated image analysis. XNAT includes methods to support each of these forms of QC. For acquisition validation, XNAT checks whether the parameters of imported DICOM files match a project-specific protocol and outputs a report detailing which tests pass and fail. The validation protocol, for example, could include a test to verify that an incoming MRI study includes a series type "MPRAGE" and three series of type "BOLD." It could subsequently test that each BOLD series has a repetition time of 2.4 seconds. The acquisition validation service is executed via XNAT's pipeline system (see "Automation" section, below). Manual image review, done soon after acquisition, is the most common and essential QC procedure. Within XNAT, a number of manual review forms have been developed to enable various levels of manual review. Typical review criteria include head positioning, motion artifacts, other artifacts, and image contrast, and may be scored either on a pass/fail or a multilevel (e.g., poor/ok/good/excellent) scale. Automated image analysis provides a quantitative approach to quality review. Using XNAT's pipeline system, specific processing routines can be executed, and quantitative metrics extracted. These quantitative metrics can then be presented in a report and compared against expected ranges or distributions generated from prior acquisitions.

All three of these QC methods are used by the HCP. Acquisition validation runs immediately after the scans are imported and checks over 20 parameters. A research analyst then reviews the structural images and enters scores into the HCP XNAT system. Meanwhile, automated QC pipelines execute to generate quantitative metrics, including temporal signal-to-noise ratio, motion displacement, and DVARS. All of these processes complete within hours of acquisition, and if key scans are deemed to be of insufficient quality, the subject is typically rescanned.

Automation

In order to prepare neuroimaging data for analysis, the acquired scans are typically run through a series of processing steps, including inhomogeneity correction, cross-modal coregistration, and denoising. A handful of widely used software packages and many in-house software libraries are available for accomplishing these tasks. While installing and running these packages is generally straightforward and can be facilitated by preconfigured operating systems like NeuroDebian (Halchenko & Hanke, 2012), the actual execution of such processing often entails running dozens of individual software routines, each with an expansive set of

execution options. Keeping track of the exact steps and procedures for a chosen processing strategy can be extremely challenging, yet it's critical to apply a consistent and systematic processing scheme for all data within the study. A number of software frameworks, often referred to as *pipeline systems*, are available for designing and executing repeatable image-processing procedures (Dinov et al., 2010; Gorgolewski et al., 2011; Marcus, Olsen, et al., 2007; Oinn et al., 2004). Pipeline systems, including the pipeline tools built into XNAT, typically entail a number of components. A specification component enables the sequence of processing steps to be defined and documented. An execution component applies a particular pipeline specification to one or more data sets. The execution component is often linked to a clustered computing environment (see below) to enable high-throughput execution. A logging component records error and status messages, including data-provenance information. Together, these components enable investigators to execute repeatable processing over time and across data sets. Pipeline systems can also be used to explore parameterized processing of a data set and to share data-processing schemes between groups. XNAT's pipeline service is fully integrated with XNAT's database, user interface, and web services, allowing users to link specific pipeline configurations to individual projects and to tie pipeline execution to their data in XNAT.

In order to capture the processing history of data generated by a pipeline (or manually, for that matter), *data-provenance* tools record processing details systematically, allowing users to confirm and verify their work and to report their methodologies alongside published data sets. Data-provenance structures typically describe an *entity* that is acted on, the *activity* that did the acting, and the *agent* responsible for an activity taking place. Within XNAT, provenance is recorded using a format developed by the Biomedical Information Research Network (BIRN) that details the input and output files to a processing routine (the entities), the details of the processing routine, including version information and input parameters (the activity), and the user and computing environment responsible for the execution of the routine (the agents; Gadde et al., 2012). For multi-step processing streams, such as pipelines, each entity-activity-agent combination is recorded as steps within an overall sequence. Recent efforts within the neuroinformatics community are moving towards the PROV data model (PROV-DM) developed by the W3C standards body as a universal model for documenting provenance of digital and real-world objects (W3C, 2013).

Anyone who has ever attempted to process large amounts of neuroimaging data knows how difficult it

can be to monitor multiple ongoing processes and to optimize use of available computing resources. Fortunately, software is available to cluster multiple computers into a coherent resource. With a cluster, users can submit large batches of jobs—indeed, more than can be run simultaneously—to the cluster. The cluster will queue these jobs and manage their execution in an optimized manner across the clustered computing hardware. Open Grid Engine (OGE) and related products are the most widely used tools for organizing and managing computing clusters. While some neuroimaging tools (e.g., FSL) are natively designed to distribute their processing across an OGE, more often users are faced with using OGE's command line tools for posting jobs to a processing queue and tracking their execution. XNAT's pipeline system, along with tools such as LONI Pipeline, natively supports integration with OGE. However, care must be taken in managing the actual computers that are on the cluster. Each of the compute nodes must have the required software (with matching versions) installed. It is also important that the nodes be running the same operating system versions and, ideally, be based on the same hardware platform. Without adequate management of the cluster, computational results may vary depending on the specific node on which a job runs.

Data integration

In addition to the acquired imaging data, neuroimaging studies often include various behavioral, clinical, genetic, and other measures. Many informatics systems provide mechanisms for capturing novel data. These include general-purpose electronic data capture (EDC) systems, such as REDCap (Harris et al., 2009), that provide web-based tools for creating data entry forms. EDCs typically utilize a set of generic key-value database tables that enable an open set of data to be captured. While REDCap and similar tools are powerful data entry systems, they do not support imaging data, so additional steps are needed to integrate nonimaging and imaging outcomes. Some imaging informatics systems, including XNAT, provide EDC-like functionality directly. Extensions to XNAT's data model are implemented using XML Schema, a format for defining XML-based data structures. From these schemas, XNAT automatically generates all of the software components necessary to import and utilize the new data type, including database tables, web-based reports, data entry forms, and programmatic interfaces.

XNAT provides a number of mechanisms for capturing these data into its database. The generated web-based forms can be used to directly enter the data,

either in real time by research subjects or a research assistant, or retrospectively by the research assistant transcribing from a paper form or some other external source (e.g., an electronic medical record). Alternatively, electronic data can be imported using XNAT's spreadsheet import service or its programmatic interface. The programmatic interface is particularly useful for implementing controlled extract, transform, and load (ETL) procedures, in which data are extracted from an external database, cleaned, and transformed into an XNAT-compliant format and uploaded to XNAT over the XNAT API. Within the HCP, ETL procedures are used to import data from the NIH Toolbox and University of Pennsylvania behavioral testing systems into XNAT. As an alternative to extending the XNAT database itself, behavioral and other nonimaging data may be stored in an external database and merged with imaging data as a final data set preparation stage prior to analysis. Data federation tools developed by the BIRN and others enable automated cross-system queries and integration (Bug et al., 2008; Zhang et al., 2011).

Security and privacy

A number of factors make security and privacy important components of neuroimaging informatics systems. Neuroimaging data contain anatomic information, including facial features, that may be used to identify the subject. Many studies also collect sensitive information—psychiatric measures, drug and alcohol usage, cognitive performance measures—that subjects expect to be maintained with discretion. These issues are particularly challenging in the context of data sharing. Institutional review boards and the federal HIPAA provide direction and enforcement on how human subject data may be accessed, stored, and distributed. The core requirements include encryption, access control, and de-identification.

Encryption is required for both storage and transport of data. While most recent operating systems natively support encryption of stored data, the feature is often not enabled by default. Encryption of data transport of the network is typically achieved by using network protocols that support secure socket layer protocols, such as HTTPS and secure FTP, which authenticates the data host and encrypts the data while in transit. However, it is important to note that most DICOM devices do *not* send data over encrypted protocols, so it is not advisable to send data from a DICOM device to a DICOM receiver that is located outside of the device's firewalled network. As a secure alternative, one can set up DICOM "relay" software, such

as Clinical Trial Processor (RSNA, 2013), inside the firewall to receive data over the DICOM protocol and then forward the data onto a receiver, such as an XNAT server, outside the firewall over HTTPS. Conversely, XNAT Gateway can be used to securely retrieve encrypted data from an XNAT system to DICOM devices like scanners and workstations.

Access control is typically maintained through password-protected user accounts. XNAT maintains an internal registry of users and their assigned roles on projects. Typically, users provide their credentials directly to XNAT, via a login page in XNAT's web application or embedded in the HTTP headers associated with a web service call. XNAT verifies these credentials prior to granting the user access to data. XNAT also supports authenticating users against external identity providers, such as a university personnel directory. Using this alternative mechanism, the user supplies his external credentials to XNAT, which then verifies the credentials against an interface provided by the external identity provider. Once a user is logged in, XNAT limits the user's access to the project to which the user has been explicitly granted access. Users are typically assigned to one of three default roles on a project: owners have read, write, edit, and delete privileges on project data and can alter other users' roles; members have read, write, and edit privileges; and collaborators have read-only privileges. Custom roles can also be created to provide more fine-grained access control. An MRI technician, for example, could be granted member privileges on a project's MRI data, allowing him to upload MRI studies while being denied access to the project's clinical and other data. In a typical configuration, a new project can be created in an XNAT deployment by any authorized user of that system. The user can grant other users access to the project and invite new individuals to become users in the system. At any time, project owners can view who has access to a project's data, their level of access, and a brief history of their access.

De-identification refers to removing information that may be used to identify human subjects, including the 18 HIPAA-defined identifiers—names, dates, phone numbers, social security numbers, and so on (US Department of Health and Human Services, 2013). Similarly, anonymization refers to removal of identifying information as well as any codes that would allow linking back to identified information. Data distributed beyond the immediate research team must be de-identified. Even within a research team, data are often de-identified to minimize the risk of accidental breaches of subject privacy. Data shared openly must be fully anonymized, which can be a difficult task to accomplish

with neuroimaging studies. XNAT provides a number of tools to assist in de-identification and anonymization. It includes a DICOM editing language and processor that can be used to remove or replace metadata fields in DICOM that may contain identifying information. All imported DICOM data is automatically processed following site-wide and project-specific DICOM edit scripts. When DICOM files are initially received, either directly over the DICOM receiver or via the HTTP interfaces, a site-wide script is applied. The default site-wide script provided with XNAT replaces the patient ID and patient name fields with XNAT session and subject identifiers. The default scripts can be edited or replaced by system administrators to enforce local privacy rules. As the data are being archived into a specific project, a project-specific script is executed to apply additional rules. If the data are subsequently moved to a different project, the new project's script is applied. Project managers can configure the project-specific scripts to provide little or no way modification, to comply with the DICOM standard for de-identification (DICOM Standards Committee, Working Group 18 Clinical Trials, 2011), or to execute a custom de-identification scheme.

A final aspect of de-identification is the necessity of removing identifying information in the actual image data. Potentially identifying facial features, including ears, must be obscured, while leaving necessary features (brain tissue) fully intact. Common procedures typically first register the scan to a target image, which provides a reference for where identifying features are positioned. An algorithm is then executed to remove or obscure these features, and the resulting image is then transformed back to the original subject space. One widely used approach uses a brain mask or tissue-segmentation algorithm to locate brain, face, and other structures, then blacks out or otherwise obscures areas outside the desired regions (Bischoff-Grethe et al., 2007). These methods are generally effective at rendering anatomic features unrecognizable, while preserving the cranial vault, but often require supervision to ensure that brain tissue is not impacted and may be affected by variation in diagnoses, age groups, MR field inhomogeneity, and other subject- and acquisition-specific variability (Fennema-Notestine et al., 2006). In addition, downstream algorithms may be disrupted by edge artifacts introduced by the defacing. Many projects have therefore moved to approaches that minimally alter the image. The method used by the HCP, for example, identifies the surface containing the face and the ears and then runs an irreversible blurring algorithm on that surface only to reduce anatomic recognizability (Milchenko & Marcus, 2013).

Data sharing

The benefits of open sharing have been well articulated by many advocates of open science (Poline et al., 2012). Inspired by this movement, many investigators are choosing to voluntarily and enthusiastically share their data. Others are mandated to share data by funding agencies that desire to see their investments put to broader use and by journal publishers who believe papers are best understood when paired with the data used to produce them. Within neuroimaging, projects such as the Alzheimer's Disease Neuroimaging Initiative (Jack et al., 2008), Open Access Series of Structural Images (Marcus, Wang, et al., 2007), the HCP (Van Essen et al., 2013), and the 1000 Functional Connectomes (Biswal et al., 2010) have demonstrated that open sharing of high-quality, well-documented data sets is beneficial to both producers and consumers of data. These projects have helped establish best practices for addressing the human-subject regulations and logistical complications that have long been roadblocks to sharing (Mennes, Biswal, Xavier Castellanos, & Milham, 2013; Poline et al., 2012). In concert with these practices, informatics systems provide a technical mechanism for sharing data.

While a dedicated instance of XNAT or other informatics platforms could be deployed to share a data set, a number of organizations have set up public sites for sharing neuroimaging data. These include the International Neuroinformatics Coordinating Facility, the Neuroimaging Informatics Tools Resource Clearinghouse (NITRC), and the Open fMRI project. Using these sites to share data has a number of advantages over operating a dedicated standalone site. For one, these organizations have worked out the technical issues and absorbed the costs of producing and maintaining the required computing, networking, and security infrastructure. Further, as the hosts of multiple data sets, they also have broader visibility within the community. The International Neuroinformatics Coordinating Facility Dataspace, which layers security, access control, and high-speed data, transfers on top of a general file-sharing service. Importantly, the service allows sharing with select collaborators or with the open community. The NITRC Image Repository builds on XNAT to provide a focused neuroimage data-sharing environment that is connected to the NITRC Computational Environment, an Amazon Cloud-based platform for executing processing and analysis routines on shared data. The Open fMRI site, in particular, is a model for open-access data sharing (Poldrack et al., 2013). The operators enforce a careful curation process for ensuring the integrity of the data that includes manual review of defaced images,

quality control metrics, and processing output. The site has a clean, ergonomic user interface and clearly documents their standard data file organization and naming conventions and the provenance of hosted data sets. Finally, its data sets are typically distributed under the unrestrictive Public Domain Dedication and License version 1.0, which allows consumers of the data to redistribute and reprocess the data in innovative ways.

REFERENCES

- BARCH, D. M., BURGESS, G. C., HARMS, M. P., PETERSEN, S. E., SCHLAGGAR, B. L., CORBETTA, M., ... VAN ESSEN, D. (2013). Function in the human connectome: Task-fMRI and individual differences in behavior. *NeuroImage*, 80, 169–189.
- BISCHOFF-GRETHER, A., OZYURT, I. B., BUSA, E., QUINN, B. T., FENNEMA-NOTESTINE, C., CLARK, C. P., ... FISCHL, B. (2007). A technique for the deidentification of structural brain MR images. *Hum Brain Mapp*, 28(9), 892–903.
- BISWAL, B. B., MENNES, M., ZUO, X. N., GOHEL, S., KELLY, C., SMITH, S. M., ... MILHAM, M. P. (2010). Toward discovery science of human brain function. *Proc Natl Acad Sci USA*, 107(10), 4734–4739.
- BUG, W., ASTAHKOV, V., BOLINE, J., FENNEMA-NOTESTINE, C., GRETHE, J. S., GUPTA, A., ... MARTONE, M. E. (2008). Data federation in the biomedical informatics research network: Tools for semantic annotation and query of distributed multiscale brain data." *AMIA Annu Symp Proc*, 1220.
- CHEN, J. J., SIDDIQUI, K. M., FORT, L., MOFFITT, R., JULURU, K., KIM, W., ... SIEGEL, E. L. (2007). Observer success rates for identification of 3D surface reconstructed facial images and implications for patient privacy and security. *SPIE Proc*, 65161B–65161B–8.
- DAS, S., ZIJDENBOS, A. P., HARLAP, J., VINS, D., & EVANS, A. C. (2011). LORIS: A web-based data management system for multi-center studies. *Front Neuroinform*, 5, 37.
- DICOM Standards Committee, Working Group 18 Clinical Trials. (2011). *Digital imaging and communications in medicine*, Suppl. 142, *Clinical trial de-identification profiles*. Retrieved from ftp://medical.nema.org/medical/dicom/final/sup142_ft.pdf
- DINOV, I., LOZEV, K., PETROSYAN, P., LIU, Z., EGGERT, P., PIERCE, J., ... TOGA, A. (2010). Neuroimaging study designs, computational analyses and data provenance using the LONI pipeline. *PLoS ONE*, 5(9).
- FENNEMA-NOTESTINE, C., OZYURT, I. B., CLARK, C. P., MORRIS, S., BISCHOFF-GRETHER, A., BONDI, M. W., ... BROWN, G. G. (2006). Quantitative evaluation of automated skull-stripping methods applied to contemporary and legacy images: Effects of diagnosis, bias correction, and slice location. *Hum Brain Mapp*, 27(2), 99–113.
- FIELDING, R. T., & TAYLOR, R. N. (2002). Principled design of the modern web architecture. *ACM Trans Internet Technol*, 2(2), 115–150.
- GADDE, S., AUCCOIN, N., GRETHE, J. S., KEATOR, D. B., MARCUS, D. S., & PIEPER, S. (2012). XCEDE: An extensible schema for biomedical data. *Neuroinformatics*, 10(1), 19–32.
- GLASSER, M. F., SOTIROPOULOS, S. N., WILSON, J. A., COALSON, T. S., FISCHL, B., ANDERSSON, J. L., ... VAN ESSEN, D. C. (2013). The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, 80, 105–124.