2

The Electronic Data Capture (EDC) System used in this study is called Labmatrix, developed by BioFortis. Labmatrix is agnostic to research domain ans is designed to manage data in a secure, centralized repository. The data servers are housed within NHGRI's secure network, behind NIH's firewall, so that only users with an NIH account, within the NIH network or logged in through the VPN, can access it. Labmatrix provides comprehensive integration with other applications and database systems, e.g. electronic medical records, clinical trial management systems, LIMS, and other research applications such as genomic or proteomic databases. This enables flexible user-driven electronic data capture forms without additional programming or customization. Labmatrix also provides flexible and fine-grained user access control and requires minimal training. It conforms to regulatory compliance standards (e.g. HIPAA, 21CFR.11) and secures PHI. Labmatrix also stores information regarding recordi modification, providing an audit trail and electronic signature features. It allows for data exporting as commonly used formats (e.g. CSV, XLSX) to enable data analysis in any major statistical software.

The version currently used is 6.3.3 (Revision: 23600).

3

Raw data that cannot be stored in our EDC (e.g. imaging or genomics data) are stored in the lab's network drive. Access to the drive is restricted to members of the lab through NHGRI's Information Technology team. The drive resides within NHGRI's secure network, behind NIH's firewall, so that only users with an NIH account, within the NIH network or logged in through the VPN, can access it.

Different types of data are stored in separate directories in the shared drive, indexed by randomized identifiers. The link between those IDs and patient information is stored in Labmatrix and in a hard-copy sotred in a secure location in the lab.

The NHGRI's IT team performs regular backups of the shared drive to mitigate data loss.

4.1

The data for this study is largely stored through Custom Data forms in Labmatrix. Each form can have many fields, and Labmatrix provides several different types of fields, which include: integers, decimals, dates, times, notes, Yes/No, and lists. They all share the option of whether the field is required to be filled in or not. Then, Labmatrix provides a Javascript interface that enables the input of code to check any entries. For example, valid-range checks, missing values, and formulas for automatic filling in fields based on other values can be set up this way.

4.2

Labmatrix periodically runs data integrity checks with respect to subject demographics. Those tests validate the demographics data entered in Labmatrix against the data entered in the Clinical Center database during the admissions process. Incongruencies are highlighted to be later manually fixed by researchers in the lab.

Data are exported in a monthly basis to run offline scripts that alert for multivariate and cross-module errors. Examples of such errors can be the same biosample ID assigned to different participants (which is later corrected based on hardcopies of ID assignments), date mismatch between visits and data collection, and sample history monitoring error (for example, a biosample DNA extraction date cannot happen before the sample collection date).

5.1

All data for this study, with the exception of subject demographics data, are stored in Labmatrix through Custom Data forms. Each form is designed to best store raw or metadata about the different types of information collected in each visit.

For questionnaires, each form mimics the question structure in the paper copy to facilitate data entry. Some questionnaires might have aggregated scores in the end, which is implemented in Labmatrix through formulas written in Javascript code. When the researcher finishes manually entering the data, the aggregated scores are calculated and saved with the corresponding record.

Neuropsychiatric tests are stored in a similar fashion to questionnaires, where each form stores the results of a specific test. Manual test scoring is needed, which is dependent on each test, and those results are then entered into Labmatrix.

Only metadata about neuroimaging scans is saved to Labmatrix. Those include the scan identifier, type of scan, date, scanner, and any notes. The actual neural data is stored in the lab network drive, as described in Section 3, and indexed by the identifier. Results of visual quality control for each scan are also stored in Labmatrix.

Biosamples data require a more complex form of storage in Labmatrix, but it is still implemented through a Custom Data form. We track each movement from the date when a sample is collected until genomic data is returned to us about the sample. That requires storing not only biosample type (e.g. Blood, Saliva, Blood DNA, SNP data, etc), but also the sample status (e.g. in inventory, shipped for DNA extraction, etc), physical location, and day in / day out (e.g. date of collection, and date of shipment for extraction).

Data is validated during entry through the specification of field types, required fields, and Javascript checks upon saving the data record. Multi-variate checks are performed posthoc as described in section 4.2.

5.2

Not sure what to do here... it's an ongoing process

6.1

Any person in the lab, prior to be granted permission to use Labmatrix, must be trained on how to use the EDC by a Labmatrix staff. After this general Labmatrix training, each new person also goes to a specific Labmatrix training with the data manager to ensure understanding of the specific design for data storage in the lab.

6.2

Persons in the lab, with exception of the data manager, acquire permissions to enter and query data in Labmatrix after initial training detailed in section 6.1.

6.3

Data in Labmatrix are entered by trained site personnel. For the most part, the data are based on source documentation maintained at the clinical site, either in digital or hard copy. For cases where the EDC system stores only metadata (scans and biosamples, for example), data are stored in the network drive by the data manager.

6.4

Accounts in Labmatrix have different roles. Only the data manager can remove records. Other members of the team can read data in the database, and enter new data. Labmatrix keeps an audit trail and regular backups in case the data need ti be restored to a specific time point. The background database is housed in redundant servers under the NIH firewall.

6.5

Quality control procedures are largely employed at data entry (see sections 4.1 and 4.2). Post-hoc quality control tests, mostly focused on multi-domain data, are performed monthly using R and Python scripts to ensure data consistency.

6.6

All data queries in Labmatrix are generated through a tool called Qiagram. It is a visual interface to construct a SQL query on the background, by creating set operations (intersection, union, filtering, etc) transparent to the user. Data queries can be performed by all members of the lab, as they do not alter the data in Labmatrix. The results of such queries can then be exported to CSV or Excel files to be later used in data analysis.

7

The only electronic data file stored in Labmatrix is the consent form (scanned PDF). The Custom Data form to store the file also includes fields to store participants' answers about data sharing that are part of the consent process. We rely on the security infra-structure that is part of Labmatrix and the NIH network to ensure data safety and integrity.

8

Not sure how to address this.

9

Data are exported from Labmatrix for any sort of data analysis. It supports CSV, tab-delimited, XML, and Excel formats.

Sometimes data are also exported to aid editing, as there are operations that are easier to be performed in a matrix format (i.e. in Excel), and then re-uploaded to Labmatrix. This EDC has the cabalitity of "checking out" records for editing, so that when re-uploaded to Labmatrix only the appropriate records are changed.

12

Not sure how to address this.