FBIRN Hierarchy

The human data storage hierarchy was designed to be amenable to automated processing scripts and robust enough to account for differences in storage requirements across FBIRN and MBIRN consortia. These storage hierarchies are general enough for any data that lends itself to the concepts of experiments, protocols, and data collection visits.

At the highest level of the hierarchy is the "Root Directory" this directory consists of an arbitrary project name, most often agreed upon by the collaborators, followed by two underscore characters "__" and a project number. The project number is obtained from the BIRN website when creating a new collaborative project. Just below the project directory is the "Data" directory indicating that the data contained are images and data files as opposed to word documents or discussion topics, etc. At the highest level inside

Root Directory .... Analysis

e.g. ResearchProject/Data
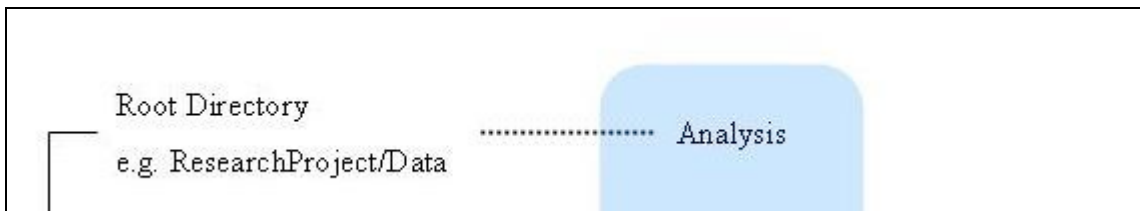
the ResearchProject__ProjectNumber/Data directory there may be an "Analysis" directory. The "Analysis" directory can contain one or more subdirectories corresponding to analyses done on many subjects, across institutions, within this project.

The second level is the subject level. The directory names are made from the BIRN ID for those subjects which are a 12 digit code; the first 4 digits are the site ID who recruited the subject followed by an 8 digit random ID. The subject level may also contain an
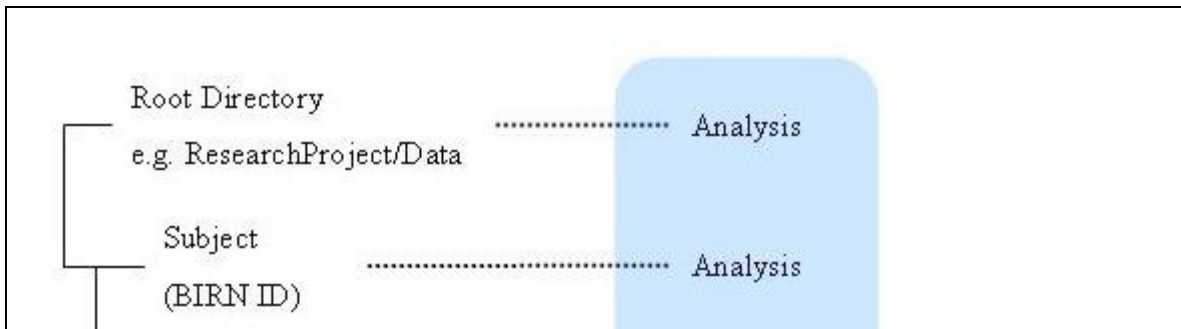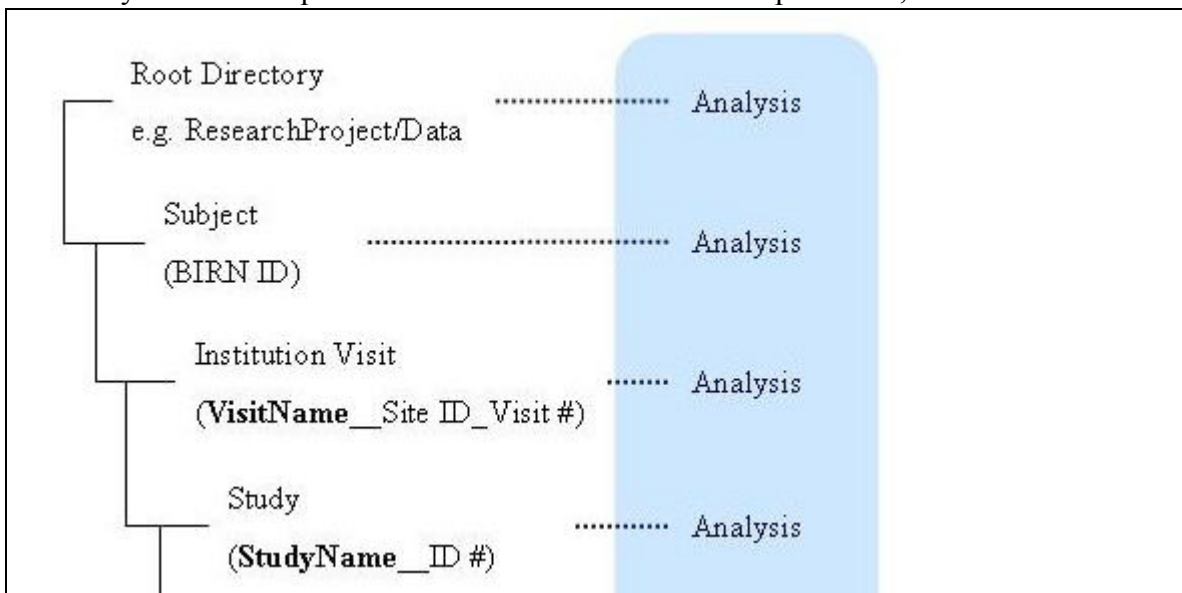
Root Directory .... Analysis

e.g. ResearchProject/Data

Subject .... Analysis

(BIRN ID)

"Analysis" directory containing one or more subdirectories corresponding to analyses done on a subject across visits. Because a single subject may visit more than one imaging site, the analyses stored at the subject level may include cross site analyses of a single subject.

The third level is the "Visit" level. Directories at this level correspond to physical visits to a location for a procedure or series of procedures (termed studies). The directory names are made up of an arbitrary visit name, followed by two underscore characters "__", followed by the site ID of the visit location, followed by two underscore characters "__", followed by the visit number at that location. The visit level may also contain an "Analysis" directory containing one or more subdirectories corresponding to analyses

done on a single subject, within one visit, but across multiple studies for that visit (see below for study definition).
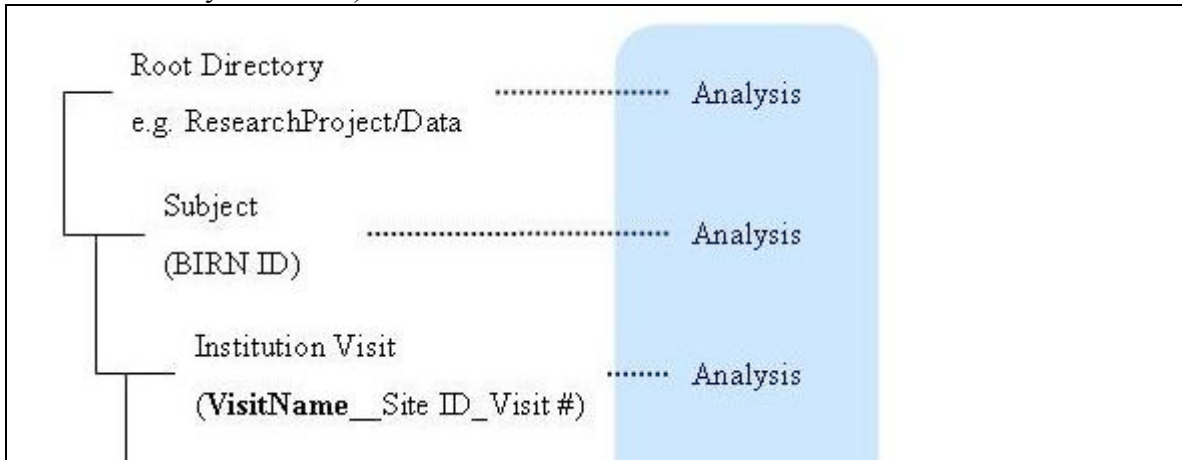

The fourth level is the "Study" level. A study is grouping of one or more procedures done during a physical visit to a site. Examples include a PET scan, MRI scan, and neuropsychological testing done during one physical site visit constituting three separate studies. The study directory name is composed of an arbitrary study name, followed by two underscores "__", followed by the study ID. The study ID should be the order which the study data was acquired within that visit. For the example above, if the PET scan was

acquired first it would get study ID 1, followed by the MRI getting study ID 2, etc. The study level may also contain an "Analysis" directory containing one or more subdirectories corresponding to analyses done on a single subject, within one visit, within one study but across multiple series in that study.

The fifth level is the series level. For imaging experiments there are generally more than one set of images collected corresponding to different pulse sequences (T1, T2, PD) or behavioral paradigms (SIRP, auditory oddball) this storage level provides the constructs to organize such data. The directory naming is made up of an arbitrary series name, followed by two underscores "__", followed by a series ID (optional). The series ID can be the series ID assigned by your scanning equipment (if desired) providing a lifeline to

Directory Hierarchy
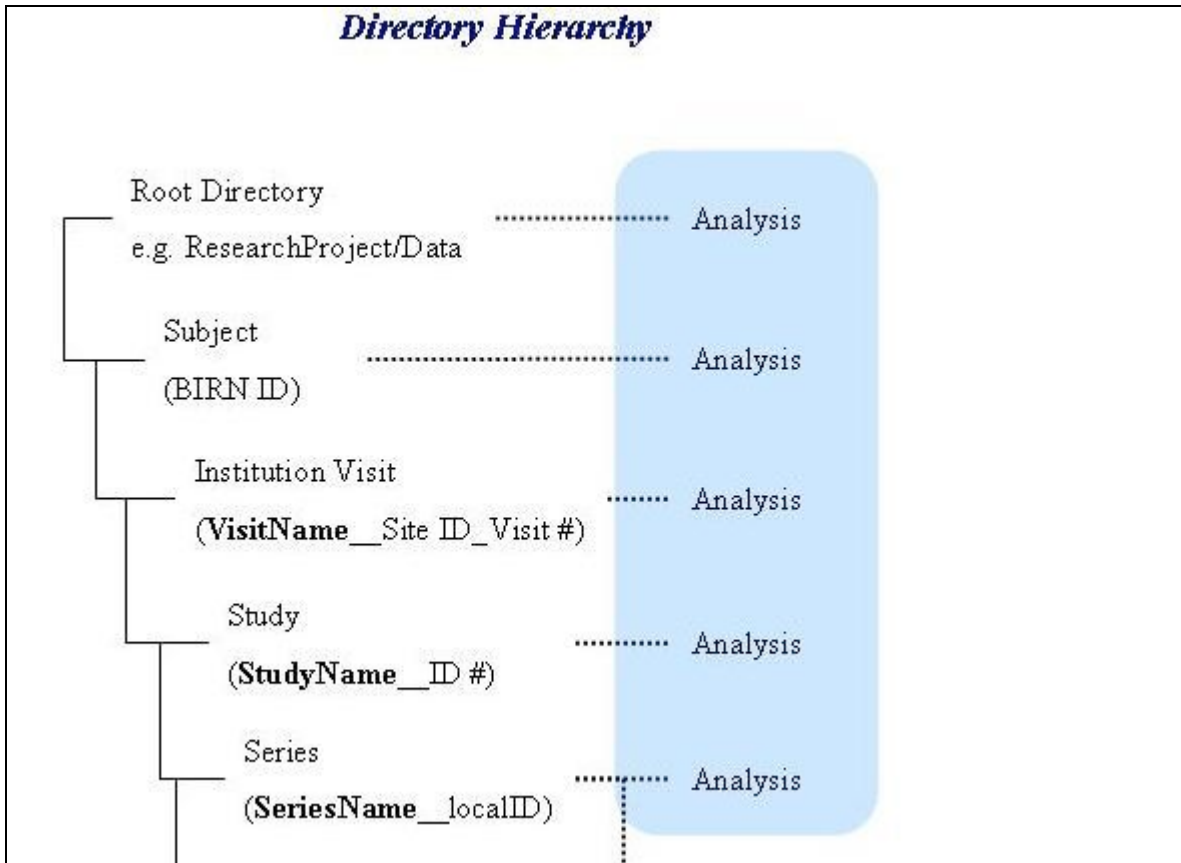
the original data on your imaging system or could be left out if not necessary. The series level may also contain an "Analysis" directory containing one or more subdirectories corresponding to analyses done on a single subject, within one visit, within one study, within one series. This analysis is equivalent to a first level analysis in fMRI studies.

The sixth, seventh, and eight levels of the hierarchy provide a mechanism for storing the native data from a data collection system. The sixth level directory name, "Native" indicates that what is below represents the reconstructed but otherwise unedited native data. The seventh level provides support for snapshots of the data. The snapshot directory naming convention is an arbitrary name, followed by two underscores "__", followed by the incremental snapshot number. Snapshots provide a mechanism for improvements in reconstruction algorithms or other such processing algorithms. For example, we might have a PET study reconstructed with filtered-back-projection algorithm stored in Native/FBKP__0001 directory indicating this was the first copy of the native data. Then we might find that some believe OSEM reconstruction is better and store the data reconstructed with that algorithm in Native/OPOSEM3D__0002, etc. The

eighth and final level provides a method of storing different data formats within a single snapshot.

Directory Hierarchy

Root Directory
e.g. ResearchProject/Data .... Analysis

Subject
(BIRN ID) .... Analysis

Institution Visit
(**VisitName**__Site ID_Visit #) .... Analysis

Study
(**StudyName**__ID #) .... Analysis

Series
(**SeriesName**__localID) .... Analysis

Native .... Analysis

Snapshot 1
DICOM AFNI
Analyze • • •
Snapshot N
(**VerName**__VerID)

Snapshot 1
Analysis Sub Tree
Snapshot N