Neuro Data Platform for Neuroscientist

Presented by: Nurul Hashimah Ahamed Hassain Malim, PhD

5th October 2018

Kami Memimpin *We Lead*

Problem Statement

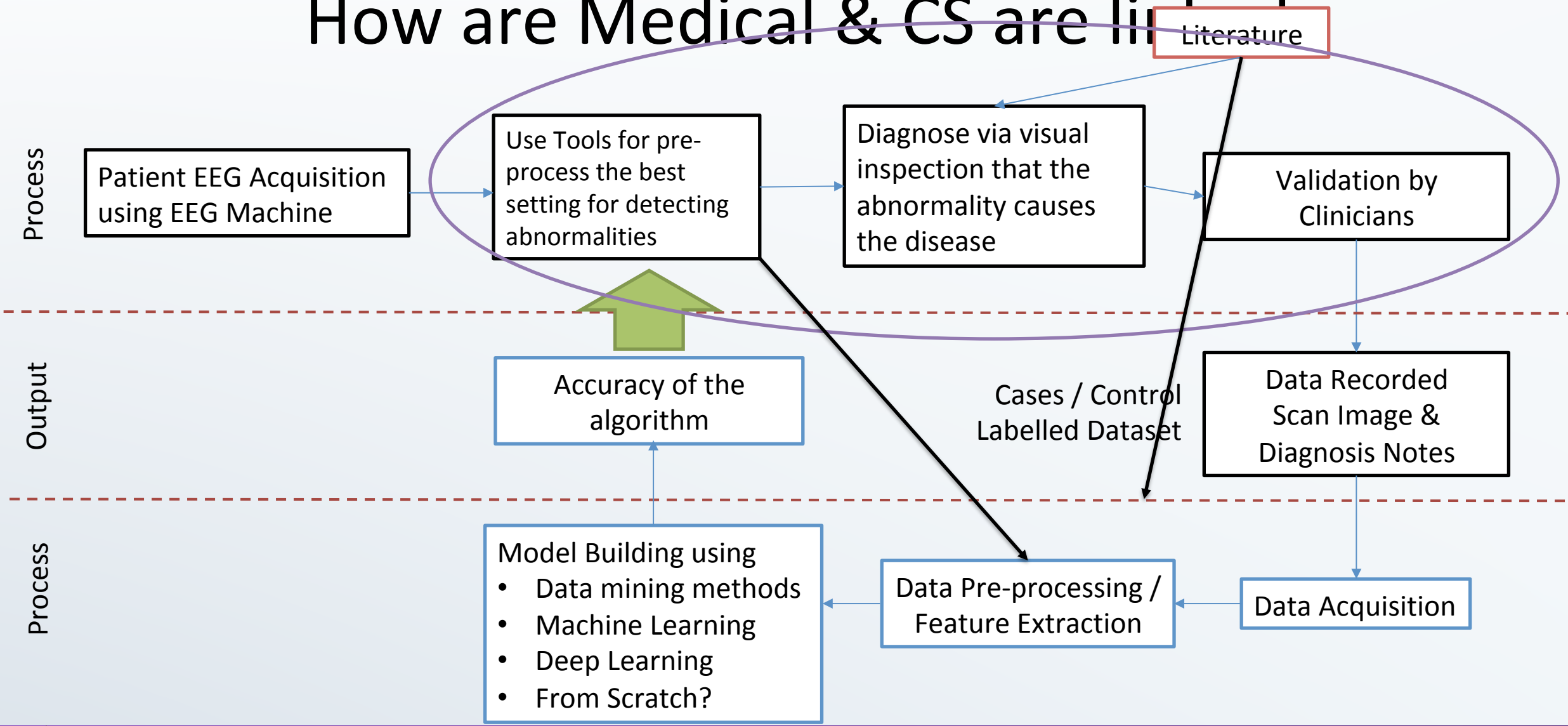
- March 2017 - CS USM received a request from P3 Neuro (now part of Department of Neurosciences) via Dato' Prof Jafri Malin to assist them on making full use of their "abandoned" data.

- 14 Aug 2017 Discussion (P3 Neuro and Nurul):
  - Their ultimate aim was to use the data for analysis but their condition did not permit them to do this yet. Plus there are differences in the context of the term analysis.

Context Differences

- Definition of analyses between Medical and computer science people are not similar
- For medical people: Medical/Imaging Analysis – using machines software to perform various analysis (coupled with their domain knowledge) to DISCOVER patient's condition – all these are recorded as DIAGNOSIS => Medical people are associating scan types to confirmed medical condition
- Computer science – Ours are retrospective analyses to uncover patterns of incident and predict what will happen – We USE the DIAGNOSED EEG/MRI and learn by using our algorithms not tools to find patterns shown by the dataset we are working on.

How are Medical & CS are li



But …

- Its not about analysing (computer science style) it yet!
- Its about preparing the data to make it ready for researchers to use for their analyses
- Emerge as Data Curation Pipeline (Advanced Data Pre-processing pipeline) for P3Neuro Data repository (hence repository management system).

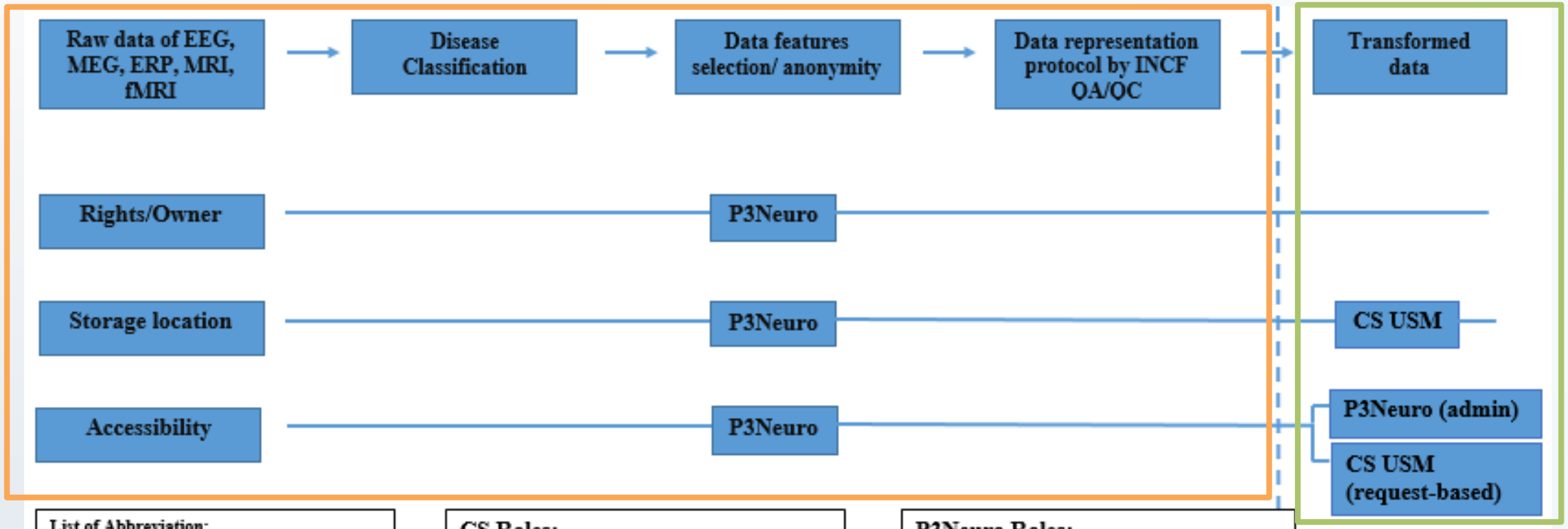
Motivation

- Why do we need curation pipeline?
  - To ease data management in the repository – so that data are
    - Free from errors (metadata) and privacy issues
    - Classified in a meaningful way for other researchers usage
    - Standardized in a general format but still maintaining its usefulness to researchers
- Why do we need repository management system?
  - As medium for curated data entry by P3Neuro and data application by researchers
  - To ease the control mechanism on the dissemination of these data

Data curation tasks

- Dissease classification – To associate each data file to specific diseases (could be one or more)
- Data features selection – To select only important attributes as metadata for storage. This is to save storage (do not keep unused/unimportant information)
- Data anonymosity – To overcome the privacy (PDPA) issues, all patient data will be made anonymous
- Data representation based on INCF protocol – To standardized data format according to the international standard

Data Curation Pipeline



List of Abbreviation:

EEG: Electroencephalogram
MEG: Magnetoencephalography
MRI: Magnetic Resonance Imaging
fMRI: Functional Magnetic Resonance Imaging
CS: School of Computer Sciences
P3Neuro: Center for Neuroscience Services & Research

CS Roles:

1) Two domain experts of medical imaging to do accuracy inspection of the task
2) To provide server storage at CS
3) To develop system to facilitate data sharing mechanism
4) To provide continous support of the system

P3Neuro Roles:

1) Three research officer to implement Task #1
2) P3Neuro will facilitate the accuracy curation
3) To manage the system
4) To update the system with the curated data

Challenges

- At first we thought that we could directly automate the curation process but after spending a few months asking people here and there for assistance and guidance, we decided to do it manually
- Two P3Neuro staffs were assigned to do this – Pn. Aida and Pn. Faten – they will explain in detail later
- PPSKOMP (Dr. Nurul) focused on repository management system – eventhough there is no indication as yet where would all these data be hosted.

Data Curation Pipeline

Progress As At 12/7/2018

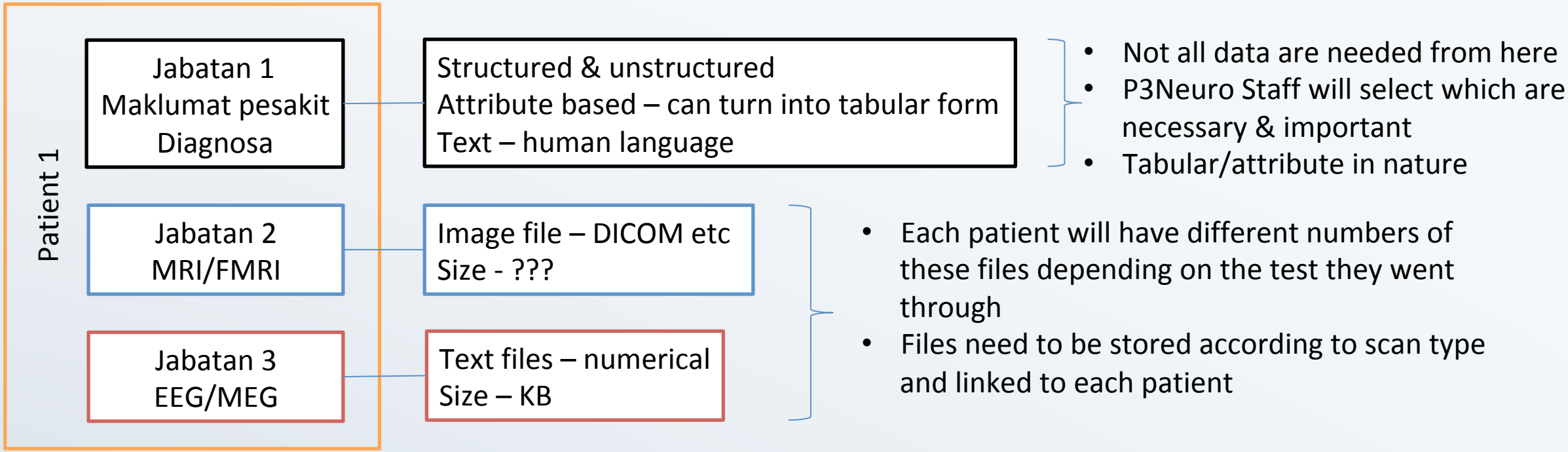
	Raw data Collection (%)	Disease Classification (%)	Data		No Of Data
			features selection (%)	anonymity(%)	
EEG	100	25	50	25	4373
MEG	100	75	100	100	190
MRI	15	0	15	0	9490
fMRI	100	100	100	100	19

Neuro repository & management system

- Users: Admin (P3Neuro) and researchers (public)
- Main Functions:
  - To store patient data
  - To allow data entry/upload, data request review and grant request by admin
  - To allow data request and download by researchers

Understanding Neuro Data

The neuro data that we would like to store are those medical scanning images of each patient.



- Not all data are needed from here
- P3Neuro Staff will select which are necessary & important
- Tabular/attribute in nature

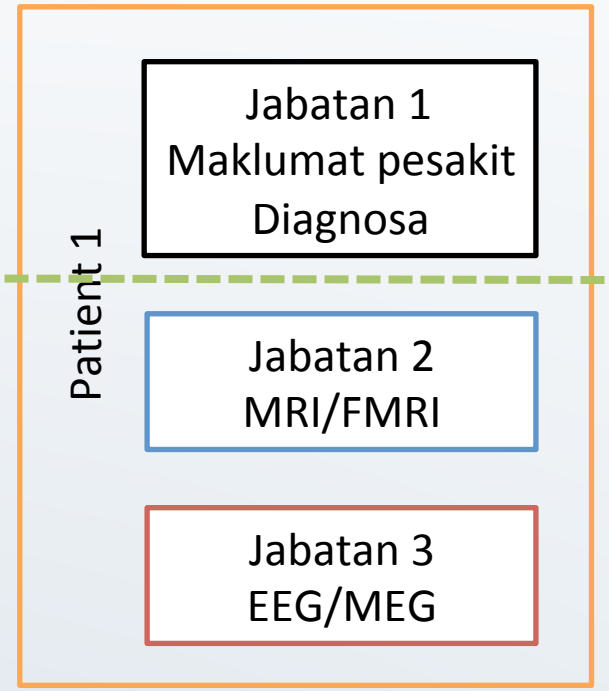
- Each patient will have different numbers of these files depending on the test they went through
- Files need to be stored according to scan type and linked to each patient

Database Management System (DBMS)

- We first need to identify what type of DBMS that is best suited for this case
- Although there is not much attribute (patient information) to be stored in the database, the instances (number of patients) are continuously growing.
- Since we are dealing with evergrowing data, we chose to use JSON as opposed to the usual RDBMS such as Mysql/Oracle

- **Why JSON**?
  - JavaScript Object Notation is a lightweight data-interchange format.
  - It is easy for humans to read and write. It is easy for machines to parse and generate.
  - Stored in a flat file => Scalable => Fast to traverse
- **Why not MySQL/Oracle?**
  - MySQL – not scalable i.e. cannot support ever growing data
  - ORACLE – licensed – and we do not need too much of the relational features. We need storage and organization of the files instead!

DBMS + Folder manageent for Neuro repository



DBMS – JSON

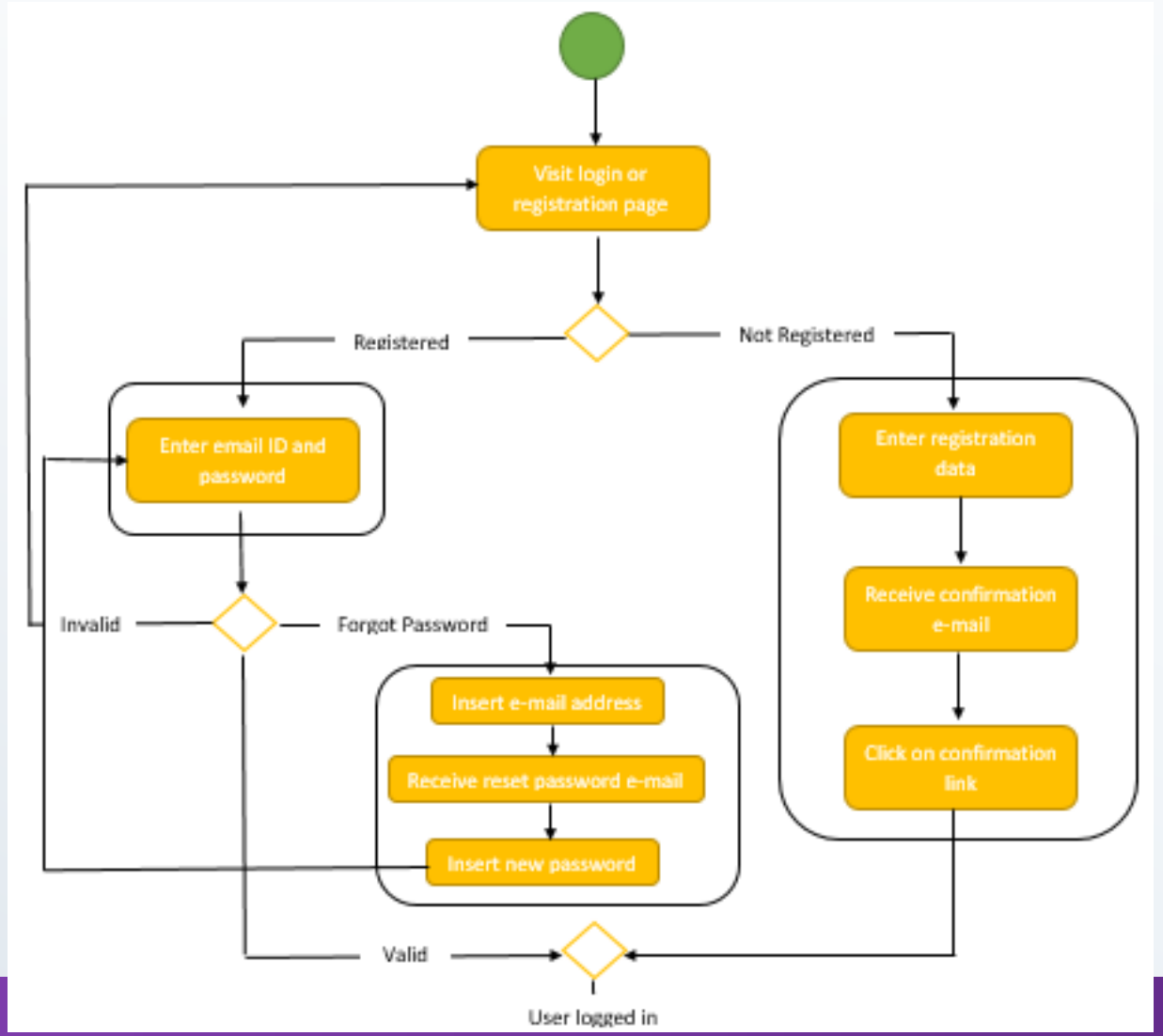
Store only attributes needed to descibe the files i.e. Patient diagnosis etc

Folder system

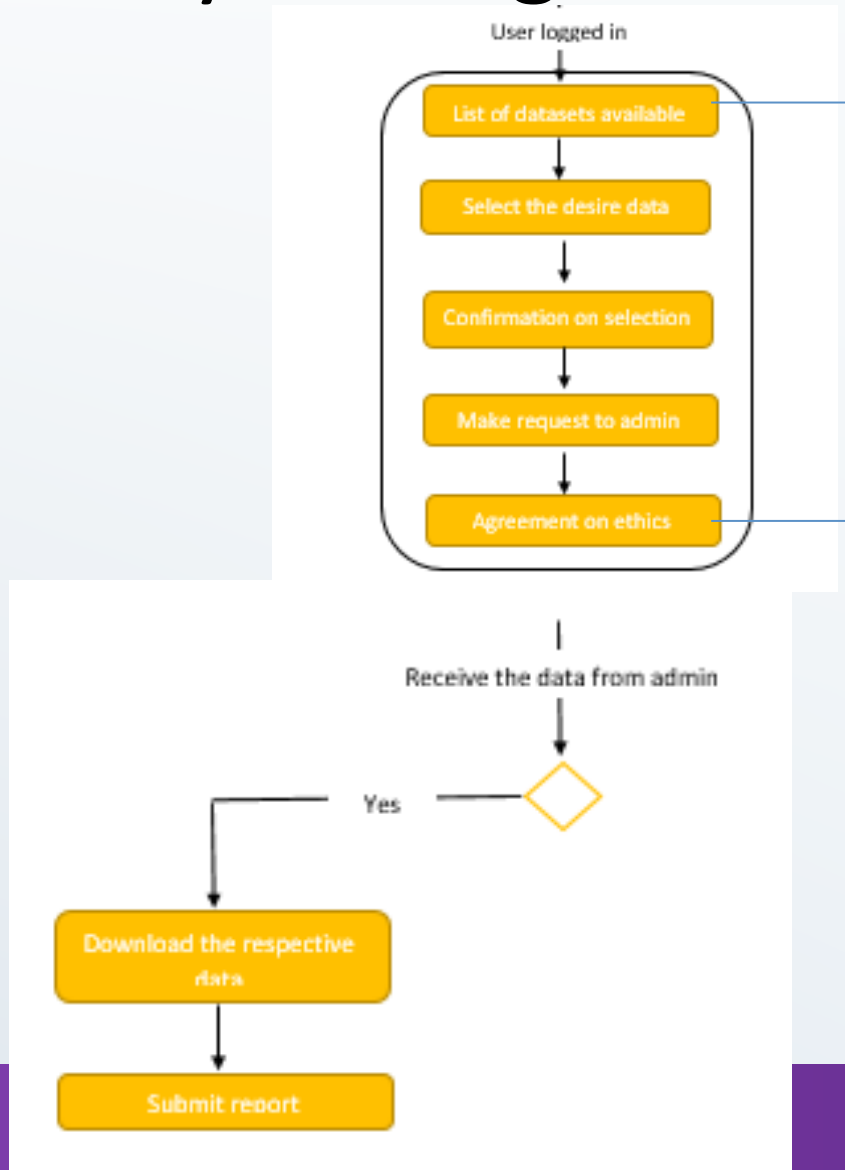
Folder will be created for each patient and subfolders of EEG / MEG / MRI etc will be created to store the image files

Neuro Repository Management System - Admin

Neuro Repository Management System - Researchers



Neuro Repository Management System - Researchers



Diseases

- Accept Terms and Condition (Data Protection Plan)
- Insert ethics approval number

Whats next

- Completing the repository system
- Venturing into the second phase – analytics for diagnosis – upon ethical clearance

EEG Team	MRI Team
AP Putra Sumari	AP Bahari Belaton
Dr. Nur Syibrah	Dr. Anusha Achuthan
Dr. Nurul Hashimah	Dr. Nur Intan Raihana

Acknowledgement

- Dato’ Prof Dr. Jafri Malin Abdullah
- Dr. Arifin Marzuki
- Prof. Dr. Ahamad Tajudin Khader
- Assoc. Prof. Putra Sumari
- Mdm Sharifah Aida Sheikh Ibrahim
- Mdm Nurfaten Hamzah
- Mr Muhammad Jaziem Mohamed Javeed

THE END