DATA TEAM INTRODUCTION

Nathaniel Edwards – Data Team Manager - nathaniel.edwards@uhnresearch.ca

Bill Li – Dados Team Lead - bill.Li@uhnresearch.ca

Justina Lam – Website Developer & Communications Manager - justina.lam@mail.utoronto.ca

UNIVERSITY OF TORONTO
FACULTY OF MEDICINE

UHN
Toronto General
Toronto Western
Princess Margaret
Toronto Rehab
COURAGE LIVES HERE

UHN Multi-Organ Transplant

DATA TEAM THEMES

- Python Programmers
- Dados Developers
- Data Engineers
- Statisticians / Data Analysts
- Website Developer / UX Developer

CURRENT STUDIES

- Explore Transplant (ETO) – Completed (http://etontario.org)
- Classical Barriers patients Study to LDKT
- PROMIS (Promis 57; Promis CAT A; Promis CAT B)
  - Expanded to cover Kidney, Kidney/Pancreas, Liver, Heart departments
- PROMIS Longtitudinal
- Expanded Barriers Patients & Non-patients
- Culture Community Survey

TEAMWORK PROJECTS

- Teamwork is a Project Management Software – a way to help us organize our projects using an online platform.
- We manage our tasks based on priority, importance, and skill sets of each individual
- Our central hub for both general topics and task specific communication
- On Teamwork, I will create a document for everyone to put in your schedule. For the work study students, I should receive your Fall/Winder time tables. If you haven't sent them to me as yet, please do. However if there are any days that you would like off from the team, please put those in the document.

DADOS

- DADOS is a web platform for electronic data collection and management in clinical and translational research, designed to bridge the gap between research and clinical care.
- It was created by integrating and enhancing two open-source web-based applications developed by Duke University, DADOS-Prospective and DADOS-Survey.
- It was designed to be fully compliant with privacy regulations (HIPAA/PIPEDA) and as such it is used to store the patients' responses to the questionnaires.
- You cannot access data from Dados from outside of UHN, unless you are using remote connection or you are using Dados External

http://dados.uhnresearch.ca/MOT/servlet/Controller

DADOS
UNIVERSITY HEALTH NETWORK
MOT
An Open Platform for Prospective Data Collection in Clinical and Translational Research

Doe, John
#999998
Logout
LAST LOGIN - 16:32 14-11-2017

SURVEYS BASELINE

In the past 7 days
I felt depressed

- Never
- Rarely
- Sometimes
- Often
- Always

PREVIOUS QUESTION | NEXT QUESTION

DADOS
UNIVERSITY HEALTH NETWORK
MOT
An Open Platform for Prospective Data Collection in Clinical and Translational Research

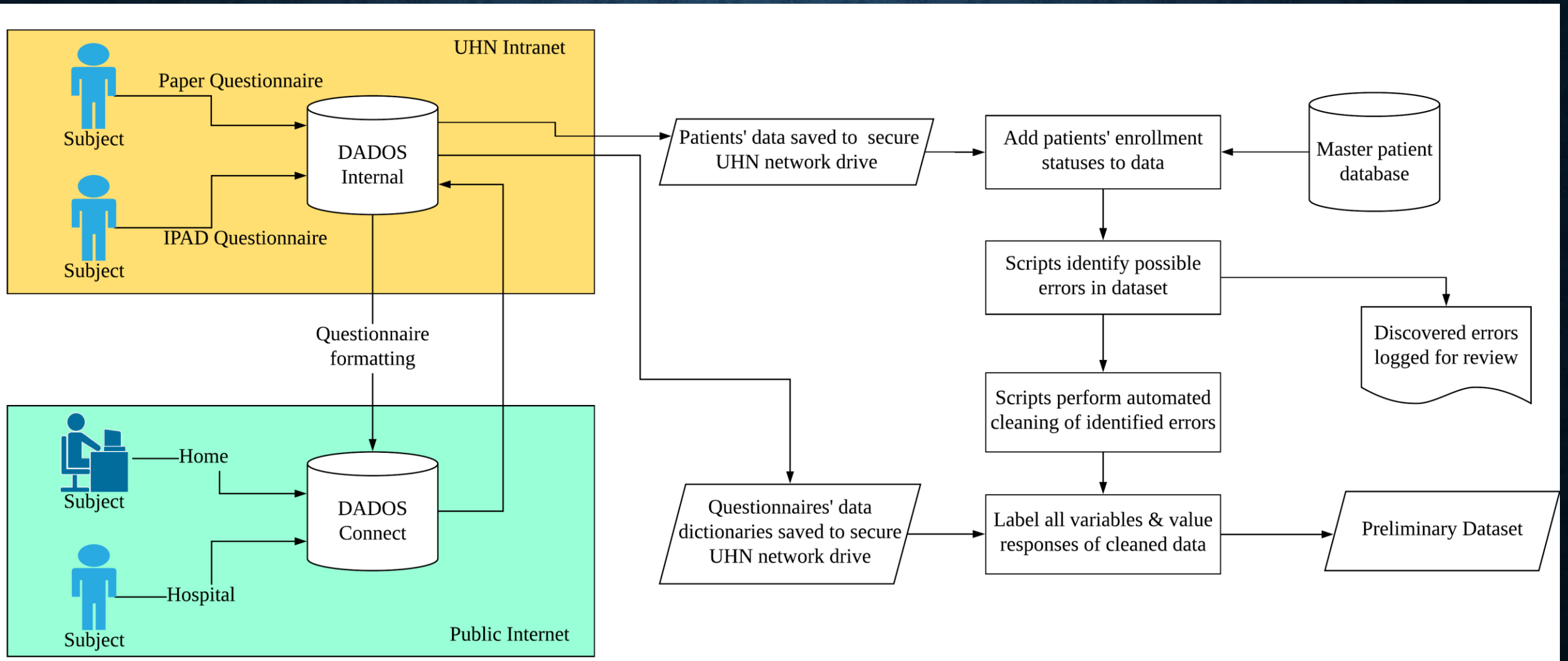
Doe, John
#999998
Logout
LAST LOGIN - 16:32 14-11-2017

SURVEYS BASELINE

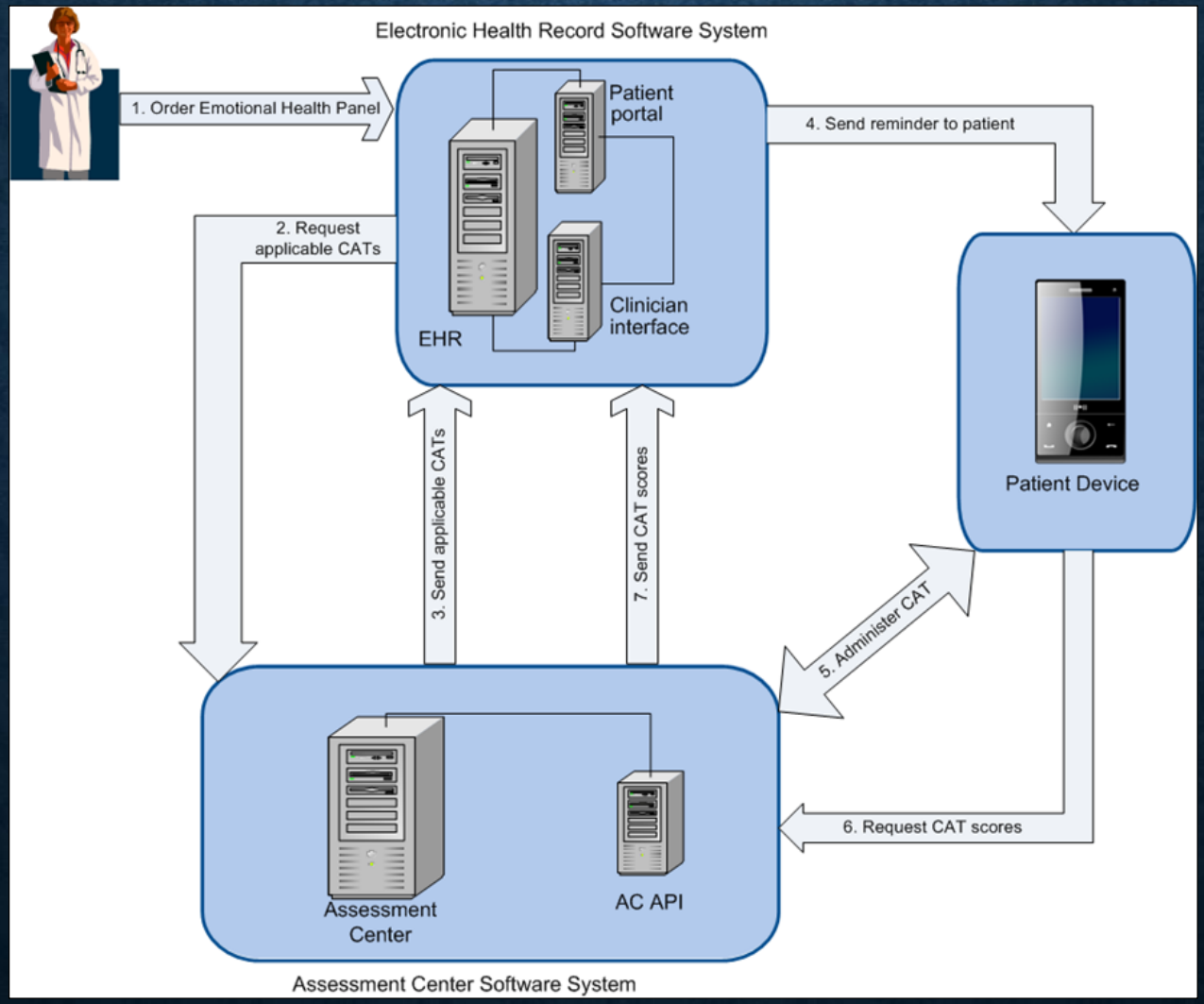
On a scale of 0 to 10, where 0 = ´No pain´ and 10 = ´Worst imaginable pain´

0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | Prefer Not to Answer | Not Applicable

PREVIOUS QUESTION | NEXT QUESTION | SAVE & EXIT


UHN Intranet
Paper Questionnaire
Subject
DADOS Internal
IPAD Questionnaire
Subject
Questionnaire formatting
Home
Subject
DADOS Connect
Hospital
Subject
Public Internet
Patients' data saved to secure UHN network drive
Add patients' enrollment statuses to data
Master patient database
Scripts identify possible errors in dataset
Discovered errors logged for review
Scripts perform automated cleaning of identified errors
Questionnaires' data dictionaries saved to secure UHN network drive
Label all variables & value responses of cleaned data
Preliminary Dataset

DADOS PROMIS CAT



CASE REPORT FORMS (CRF)

- This is not something you will need to directly work with on the data team, but it is important to understand what it is.
- When recruiters approach patients, we ask them socio-demographic questions etc. However we obtain clinical values & blood work from UHN patient records (e.g. OTTR, EPR etc).
  - Examples of these values are GFR + other lab results (hematology & chemistry) + blood type; cause of End Stage Renal Disease ESRD (e.g., Diabetes, Hypertension, Iga Nephropathy, Polycystic kidney Disease, etc.); Family history of CKD & Co-morbidities (via the Charlson Scoring System)
- These values are manually entered by recruiters into DADOS.

MICROSOFT ACCESS

- Originally designed to replace Excel tables for recording patient records.
  - Multiple users trying to enter/view/modify patient information at the same time
  - Create a structure to the data and control the data being entered
- Microsoft Access is used to store all the patient identifying information.
- Has information about patient screening, consent status, progression during each study, and status.
- Used by all research teams located in Toronto General Hospital
- Other sites (Humber & Church) are still using Excel sheets to manage their patients enrolled, declined or screened out in their studies. One of our upcoming projects includes redesigning their Excel sheets to control the data structure

EXCEL MASTER DATA DICTIONARY

- A combination of all the data dictionaries from all ongoing studies (ETO, Barriers, PROMIS, PROMIS Cat & Pilots).
- Used to link variables across studies and maintain accurate & uniform questions & answers across all studies
- All changes to variables in Dados need to be reflected here as well.
- Location:
  - U:\data\extras

WHERE DOES PYTHON COME IN? (I)

- Python is the language used to automate our data extraction
- It runs at several intervals: hourly, daily, weekly
- It extracts data from DADOS, Access, Excel files, and flat files
- Does some data cleaning
- Triggers STATA to run for scoring of medical questionnaires & unit testing

WHERE DOES PYTHON COME IN? (II)

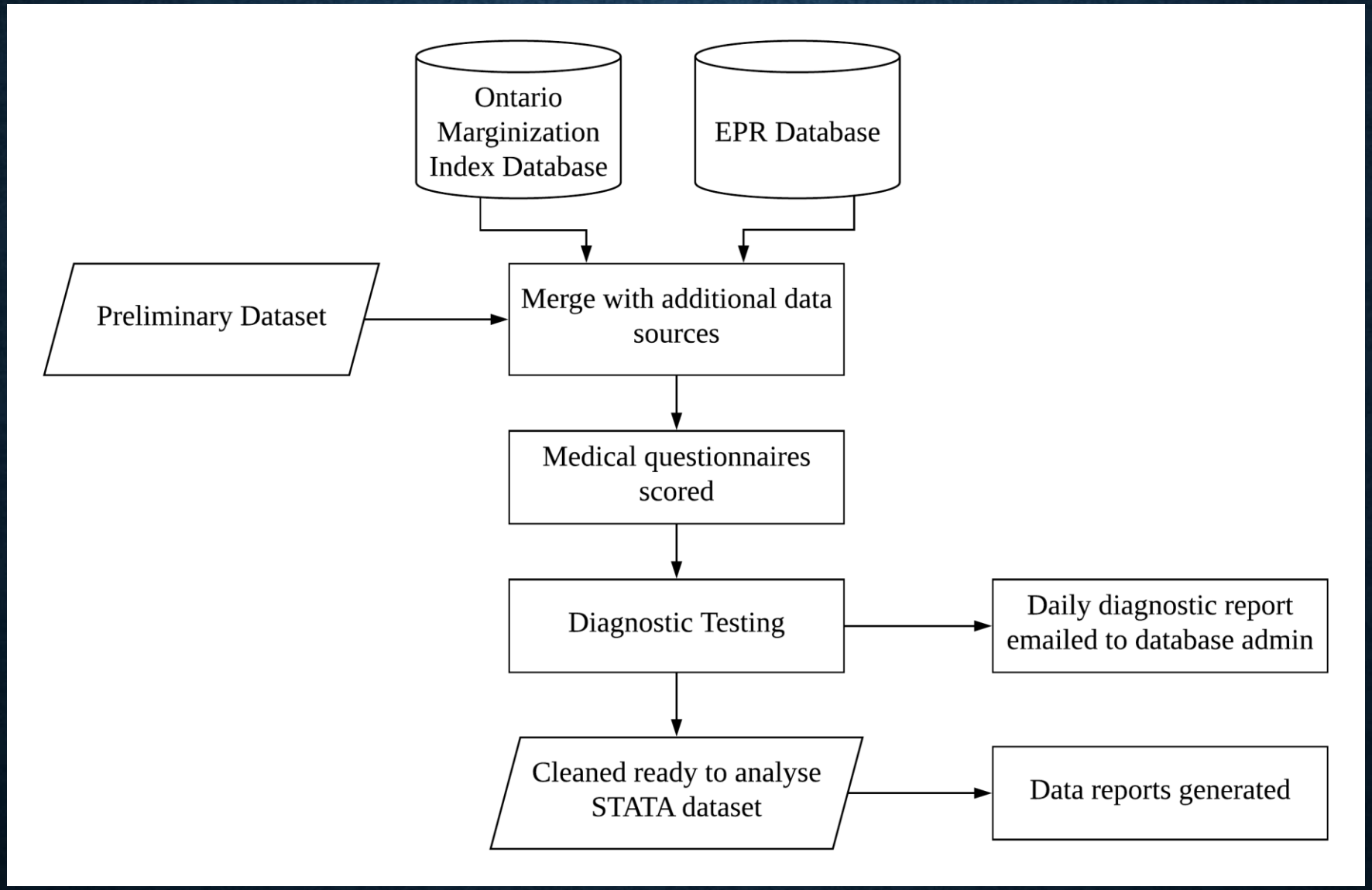
- Python scripts:
  - Python automates creating labels for variables for easier interpretation in STATA
  - It attempts to find mistakes that are missed by recruiters
  - It runs the major STATA code for generating various important variables
  - Unit testing (Diagnostic tests) to make sure the data is generated as expected
  - Sends out emails and reports to team leads based on errors encountered

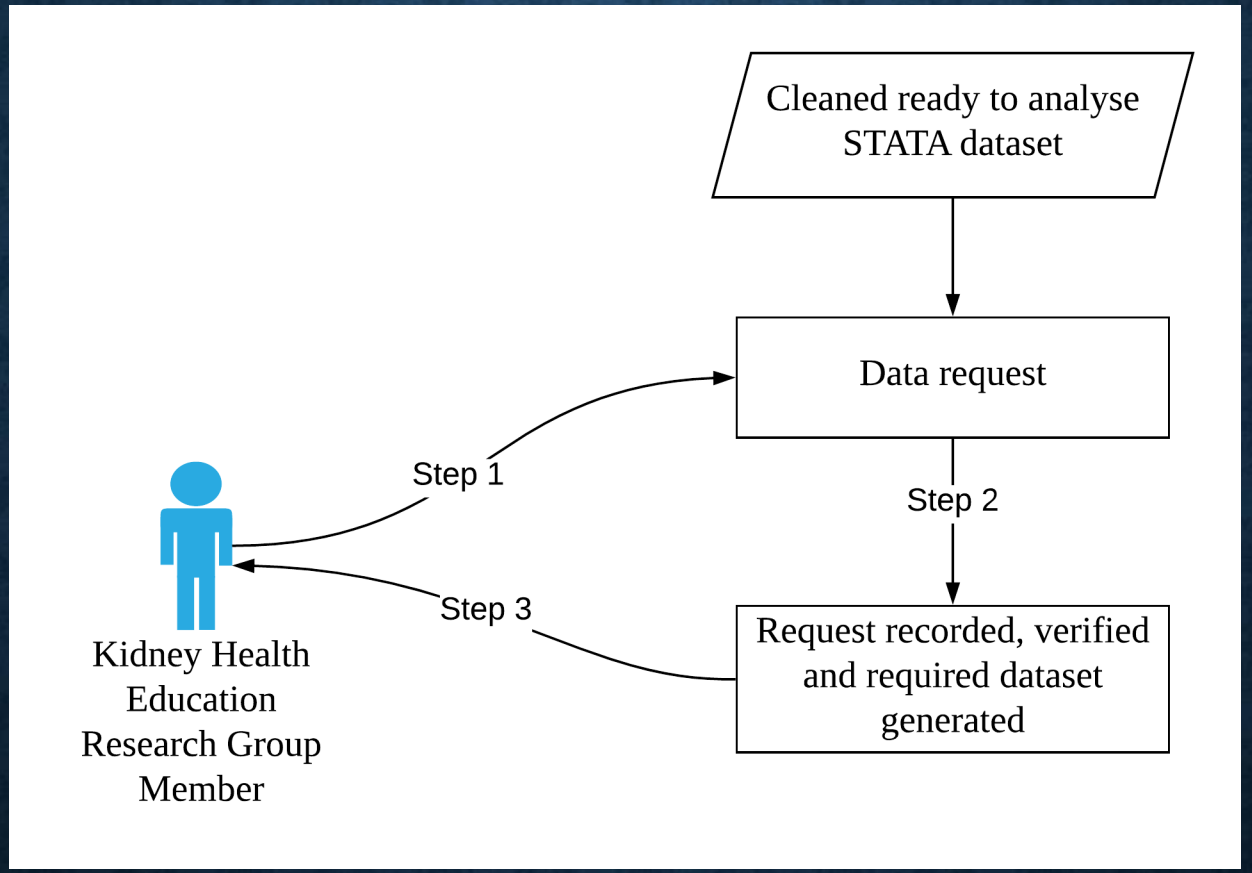
DETAILED EXPLANATION (I)

- Everyday at 12:10 am, a python program is triggered from the network drive to run automatically using Windows Scheduler. It logs into dados using one of our UHN user profiles. It requests a new data dump from dados. By dados' security this dump is sent to the user account's email address. The program logs into that email address and downloads the requested data dump to the network drive. At the same time, Python grabs all the patient information stored in the Access database, so we can identify which patients recorded in Dados are enrolled in the studies.
- It is important to remember that Dados stores the responses which our patients give to the questionnaires, but it is not used to manage the enrollment statuses of the patients. The patients' enrollment statuses is managed securely in the Access database

DETAILED EXPLANATION (II)

- Once the data is all on the network drive, the python program performs a few data cleaning steps and then triggers a STATA program to be run which performs other several processes and manipulations on the data. All questions are also medically scored in this process.
- Recruiters will have made several errors in data collection which we aim to resolve on our end. We keep track of all the questions in a master dictionary stored on the network drive and based on this, Python creates the labels for the questions and the STATA code to merges the respective questions across studies into one.
- Several diagnostic tests are also run on the data once the final STATA .dta file is prepared to ensure data quality. An email is then sent every day to our email address giving us a report of the data and which diagnostic tests fails, so we can follow up with team leads very quickly on any issues in the data.


Ontario Marginization Index Database
EPR Database
Preliminary Dataset
Merge with additional data sources
Medical questionnaires scored
Diagnostic Testing
Daily diagnostic report emailed to database admin
Cleaned ready to analyse STATA dataset
Data reports generated


Cleaned ready to analyse STATA dataset
Data request
Step 1
Step 2
Step 3
Kidney Health Education Research Group Member
Request recorded, verified and required dataset generated

QUESTIONPRO
BARRIERS EXPANDED NON PATIENTS

- Online survey platform that will be used to send out questionnaires to a larger audience. Currently we have three cultural teams which are working with their communities to reach out to the public, educate people about kidney transplantation.
  - http://nefros.net/chinese-kidney-guide-project/
  - http://nefros.net/muslim-kidney-guide-project/
  - http://nefros.net/the-south-asian-kidney-guide-project/
    - These questionnaires fall under the Barriers Expanded Non-patients research study
- Why was Dados not used for this project?
  - Dados does not allow for anonymous data collection
  - It is heavily secured and restricts access to protect patient data

WHERE DOES PYTHON COME IN? (III)

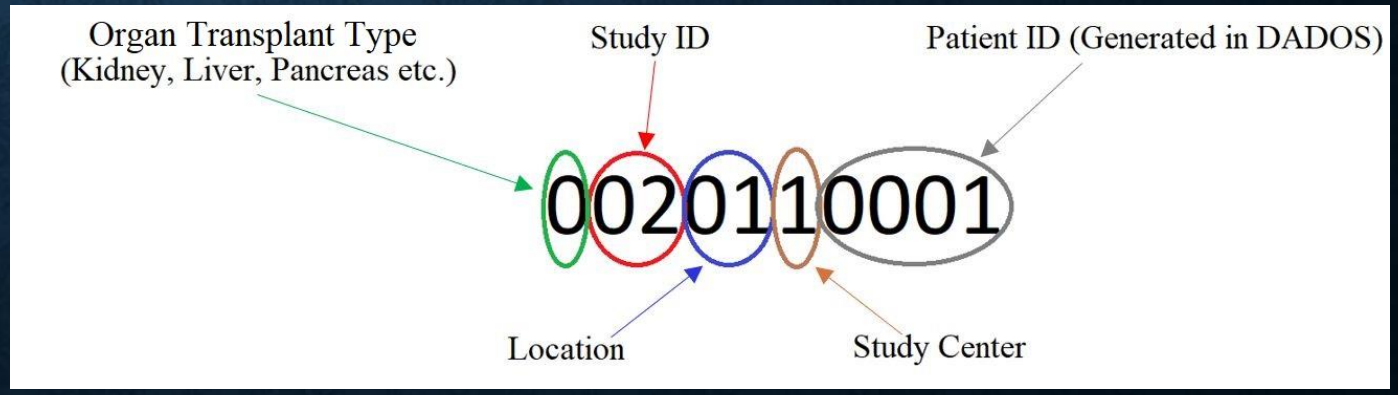
- Python is also being used to link data from QuestionPro back into Dados and into our STATA datasets for analysis

HOSPITAL MRN VS SUBJECT ID

- Every patient who checks into a hospital is assigned a unique hospital mrn which is printed on their wristband. For academic research, by the UHN's policies of protecting patients' data, we cannot reference a hospital mrn in analysis. As such all patients are assigned a unique subject id when we enroll, decline or screen them out of our research.

SUBJECT ID BREAKDOWN – 10 DIGITS

- Organ Transplant Type:
  - 0 = Kidney
  - 1 = Kidney/Pancreas
  - 2 = Liver
  - 3 = Heart
- Study ID:
  - 01 = Barriers Classical Patients
  - 02 = Barriers Pilot Study;
  - 03 = PROMIS
  - 04 = ETO
  - 05 = PROMIS Cat
  - 06 = PROMIS Cat A
  - 07 = PROMIS Cat B
  - 08 = Barriers Expanded Patients
  - 16 = Longitudinal PROMIS Cat A
  - 17 = Longitudinal PROMIS Cat B
- Location:
  - 01 = UHN
  - 02 = Humber River
  - 03 = St. Michael's
  - 04 = Sunnybrook
  - 05 = Brampton
  - 06 = Scarborough
- Study Centre:
  - 0 = Pre-transplant clinic (Heart Failure, Liver Clinic)
  - 1 = Post-transplant clinic
  - 2 = Renal management clinic
  - 3 = Hemodialysis
  - 4 = Home hemodialysis
  - 5 = Fast-track
  - 6 = In-patient 7th floor (side A and B)
  - 7 = PKD Clinic
- Patient ID (From DADOS): 001, 002 etc.



CONFIDENTIALITY OF DATA

- This topic will be covered heavily in your UHN research computing orientation. You need to attend this orientation to get set up with a UHN profile and then you will be able to have remote access

  Location: TGH
  Date: Held every Monday at NU 1N-130
  Time: 2PM

- Once you have remote access, it may be very temping to copy all the data to your personal machine and work on it there. It is very important to note that **this is a violation of UHN policy** as no confidential patient information is to be stored off site or on personal computers.
- Details of these training sessions will be confirmed by the team manager

HOW CAN YOU ACCESS THE DATA?

- Remote access – equipped with a network version of STATA (U:\Stata15)
- Come into TGH in person and connect to Dr. Mucsi's network drive on your computer.
  - Instructions for how to do this can be found on the network drive U:\Collection of Manuals and SOPs
  - It is encouraged to work offsite as there is limited space on site which is preferred to be utilized by our recruitment staff

ALL THIS DATA. WHERE IS THE ANALYSIS?

- Data analysis involves (in my opinion):
  - 80% time spent on data cleaning and management.
  - 15% of the time testing statistical models to ensure they meet the assumptions.
  - 5% of the time actually doing the statistical analysis.
- When does the analysis happen?
  - The team submits abstracts to various conferences – American Transplant Congress (ATC), Canadian Society of Transplantation (CST), European Association of Psychosomatic Medicine (EAPM) etc.
  - Researchers on the team who are working on abstracts & publications need statistical help
  - Previous analyses; papers/abstracts are stored on the network drive (U:\Conference Abstracts)
- All statistical analysis is performed using STATA

WHAT DO WE ANALYSE ANYWAY? MEDICAL QUESTIONNAIRES:

- ESAS-r Edmonton Symptom Assessment Scale
- TDMS – Transplant Decision Making Scale
- SDI - Social Difficulties Inventory
- GAD 7 - Generalized Anxiety Disorder
- WHODAS - World Health Organization Disability Assessment Schedule 2.0 etc.
- PHQ-9 Patient Health Questionnaire
- KDQOL 1.3 – Kidney Disease Quality of Life
- ISI - Insomnia Severity Index
- SLS - Short Literacy Survey
- ECR - Experiences in Close Relationships
- IIRS - Illness Intrusiveness Rating Scale
- MOS – Medical Outcomes Study Social Support
- PRQ - Patient Report Questionnaire
- FSS – Fatigue Severity Scale
- HAQ-DI - Health Assessment Questionnaire
- Global Health
- FACIT - Functional Assessment of Chronic Illness Therapy etc.

WHAT DO WE ANALYSE ANYWAY?

- You should familiarize yourself with the questions that are asked to the patients in these questionnaires so you understand what exactly is being looked at in any dataset's analysis. You should also familiarize yourself with all the abbreviations used to refer to these questionnaires.
- The master data dictionary will contain a full listing of all questions that we are. Alternatively, the individual scoring guides for these questions are located on the network drive:
  - U:\data\Questionnaires

ABBREVIATIONS TO KNOW

- EPR – Electronic Patient Record
- OTTR – Ontario Transplant Tracking Record
- CRF – Case Report Form
- ETO – Explore Transplant Ontario
- MRN – Medical Record Number
- SIMS – Shared Information Management Services
- DART – Distress Assessment & Response Tool
- REB – Research Ethics Board
- RIS – Research Information Systems

CLOSING NOTES

- Please read all documentation associated with our program and Dados
  - U:\data\documentation
  - If anything is unclear, please let us know so we can clarify
- Please read the research protocols!
  - Located on the network drive
    - U:\PROMIS-57\Research Protocols and Consent Forms\Protocol
    - U:\Explore Transplant\Protocol
    - U:\Barriers Study\Protocols
    - U:\Pilot Study\Protocols
  - Located on the website
    - http://nefros.net/members-tgh/research-protocols/

QUESTIONS?

- Email me: nathaniel.d.j.edwards@hotmail.com


"That's all Folks!"