Steps followed to get the data from the INSTA database:

1. Login onto the INSTA website
2. Go to the Patient Search tab
3. Identify unique patients
4. Use the unique URL for each patient MRNO and MR\_NO to get the source data for each patient
5. Achieve #4 by creating a VBA macro
   1. Access the source code for each patient
   2. Filter the source code and get the displayURL parameter from the Java code
   3. This parameter contains links to all the data for individual patients
   4. Save TXT files for each patient
6. Create a R program to read all the TXT files
   1. Process this data to get the patient number, visit combination
   2. Create a 1-1 match for Patient + Visit + visit date + unique number of visit to hospital
   3. This creates 4.25 lakh URLs for almost 50,000 patients in August 2017
7. Data coming from URLs, there are 20 different types of programs proving information

|  |  |  |  |
| --- | --- | --- | --- |
|  | Name of the program | | Number of URLs |
| 1 | /vitalForm/genericVitalForm.do?method | generateReport&visited | 72026 |
| 2 | /pages/GenericDocuments/GenericDocumentsPrint.do?\_method | print&forcePdf | 7238 |
| 3 | /emr/print.do?method | printConsultation&consultation\_id | 117066 |
| 4 | /emr/print.do?method | printTriageSummary&consultation\_id | 19753 |
| 5 | /dischargesummary/dischargesummaryPrint.do?\_method | print&patient\_id | 6428 |
| 6 | /Service/ServicesConductionPrint.do?\_method | print&prescription\_id | 165746 |
| 7 | /pages/DiagnosticModule/DiagReportPrint.do?\_method | printReport&reported | 29170 |
| 8 | /progress/PatientProgress.do?\_method | getPrint&mr\_no | 2623 |
| 9 | /wardactivities/DoctorOrderPrint.do?\_method | print&patientId | 2434 |
| 10 | /wardactivities/VisitSummaryRecord.do?\_method | generateReport&patient\_id | 2607 |
| 11 | /IntakeOutput/genericIntakeOutputForm.do?method | generateReport&visited | 135 |
| 12 | /dietary/DietaryMasterPrint.do?method | printPrescription&patient\_id | 7 |
| 13 | /Service/ServiceReportsPrint.do?\_method | print&forcePdf | 4 |
| 14 | /Radiology/TestDocumentsPrint.do?\_method | print&doc\_id | 84 |
| 15 | /Dietary/DietaryGenericDocumentsPrint.do?\_method | print&doc\_id | 1 |
| 16 | /MLCDocuments/MLCDocumentPrint.do?\_method | print&doc\_id | 2 |
| 17 | /wardactivities/DoctorsNotes.do?\_method | generateReport&patient\_id | 82 |
| 18 | /InitialAssessment/InitialAssessmentPrint.do?\_method | printInitialAssessment&consultation\_id | 2 |
| 19 | Discharge | *NA* | 1 |
| 20 | /GenericForms/GenericFormPrintAction.do?\_method | print&printerId | 1 |

1. The cells marked in yellow in the table above represent the pdf files not downloaded from the INSTA database.
   1. Vital form data is available in a CSV file download
   2. Lab data is available in a CSV file download
2. Use CURL command in Cygwin to extract the individual PDF files. This function works when a user is logged into INSTA system E.g.

curl 'https://182.71.223.195/instahms/Service/ServicesConductionPrint.do?\_method=print&prescription\_id=175665&printerId=2&doc\_id=175665' -H 'Accept-Encoding: gzip, deflate, br' -H 'Accept-Language: en-US,en;q=0.8' -H 'Upgrade-Insecure-Requests: 1' -H 'User-Agent: Mozilla/5.0 (Windows NT 6.3; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/60.0.3112.90 Safari/537.36' -H 'Accept: text/html,application/xhtml+xml,application/xml;q=0.9,image/webp,image/apng,\*/\*;q=0.8' -H 'Referer: https://182.71.223.195/instahms/loginForm.do' -H 'Cookie: JSESSIONID=D2D79C9FEC4DC1AB75706C93C80AE523; user\_schema@/instahms=iaim; firstMenuIndex=0; lastUser=VmluYXltYWhhamFu' -H 'Connection: keep-alive' -H 'Cache-Control: max-age=0' --compressed –insecure > FILENAME.pdf

1. Some of the filenames have “?” at the end of the file, so use DOS command to change the filename
   1. i.e. remove “?”At the end of the filename, this is done by using “rename” command as follows: rename \*.pdf? \*.pdf.
   2. The ? at the end of the files make these files unavailable for additional text processing
2. Use pdftotext command in Cygwin to convert the PDF files into txt files.
   1. Options used pdftotext –layout –nopgbrk provides the same data in text format and
   2. It reduces the file size significantly, e.g. 600 KB pdf files becomes 10 KB file,100 KB pdf file becomes 2-3 KB file.
   3. The following programming statements created TXT files for each patient in pat\_txts folder

vinay@viany /cygdrive/c/Users/Lucky/Documents/Hospital\_data/04\_2017\_DOWNLOAD/pat\_pdfs

$ find . -name "\*pdf"|sed 's/.pdf//g'|head -n 10|awk '{print "pdftotext -layout -nopgbrk", $1 ".pdf", "../pat\_txts/" $1 ".txt"}'|sh

1. Use these text files as input to various R programs and create source data.
   1. Various data tables to be created
2. SQLite freeware database will be used to store the data.