Data Science Africa 2017

Crowd Sourcing Clinical Data from Small Health Facilities in Kenya.

Dr Moses M Thiga
Kabarak University
Department of Computer Science and Information Technology
mthiga@kabarak.ac.ke

Introduction

The power of big data analytics can be harnessed for the practice of public health informatics if it is collected in a timely and accurate manner in order to facilitate its timely analysis, interpretation and dissemination for action.

Introduction

This is however **not yet possible in Kenya** given the lack of integrated and affordable clinic management systems for use **especially** by **small, financially and technologically challenged clinics, often times located in informal settlements and rural areas.**

Big Data?

Some Numbers … Approx 500 clinics in Nakuru county seeing at least 10 patients per day = 5,000 records x 6 days a week = 30,000 records a month
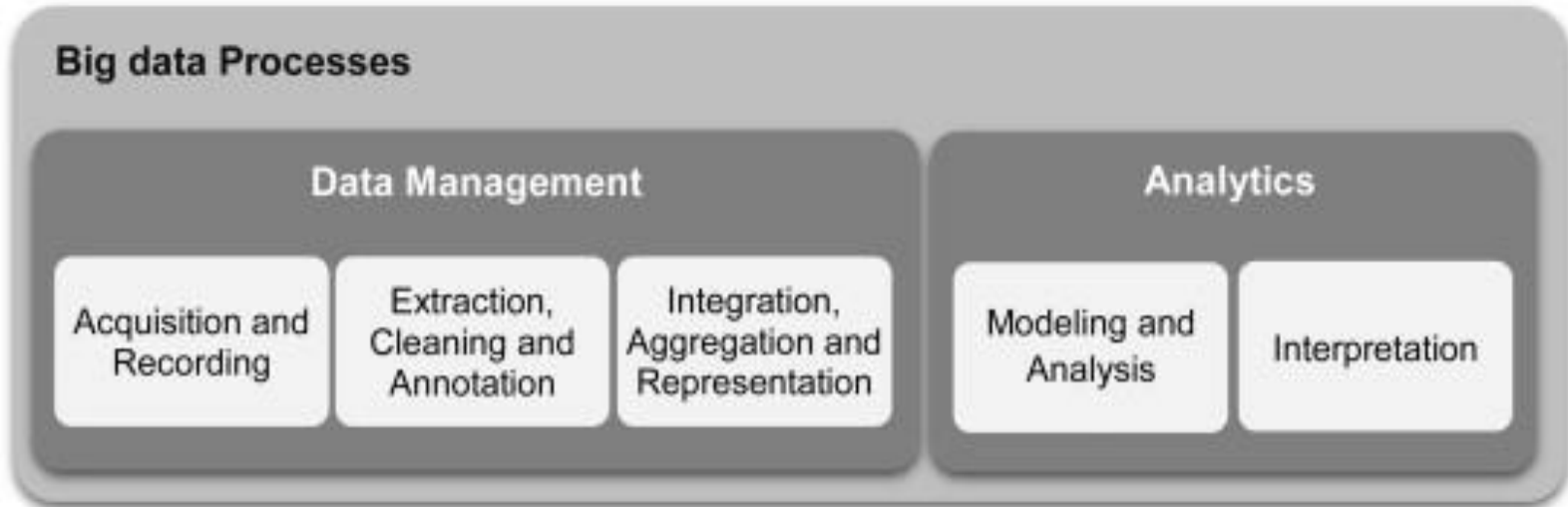
Extrapolate this to at least 20 counties → 20 x 30,000 – 600,000 records per month *if the people only make one visit in a month*

→ per year 600,000 x 2 = 1,200,000 million records

… a very very conservative estimate

The Big Data Process

Jifa and Lingling (2014)

In order to achieve effective and timely interpretation of data the data management stages must be undertaken accurately and within reasonable time limits for the ensuing information to be of value in practice.

Need for speed and accuracy

This need for speed and accuracy is generally important in all fields of application of data science but is even more critical in the practice of **public health informatics** where the timely and accurate data collection aids greatly in **biosurveillance and managing outbreaks** (American Medical Informatics Association (AMIA), 2016)

The Kenyan situation

In Kenya, the practice of public health informatics and biosurveillance as supported by big data analytics is still a dream for a number of reasons.

1) A majority of medical facilities that are automated utilize different and incompatible clinic management systems.
2) A majority of medical facilities that are (i) small and (ii) located in informal and rural settlements are not automated at all due to their inability to afford and utilize the available solutions.

The Kenyan situation

- The **Kenya Health Information System (http://hiskenya.org)** based on the popular feature rich **DHIS2** is an example of a system in use in Kenya.
- The system has been rolled out to government health facilities and is reported to be working fairly well.
- However, the lack of computers at all health facilities has led to the maintenance of hardcopy record books at small clinics.
- These hardcopy records are summarized and forwarded to district level health facilities for keying in into the system.

The Kenyan situation

This approach has a number of key shortcomings;
  1) The data captured is not real time as it has to be summarized on a monthly basis and forwarded to data entry centers. At best reports can be made available after two months from the date of a patient visit.
  2) The process of summarizing the manual records inevitable leads to data loss and introduction of errors thus making the end product less reliable.
  3) The use of the system has not trickled down to all facilities as would be desired in an ideal setup due to the investment required to make this happen.

A way forward

A suitable solution for this scenario would allow for data to be captured at source into the main system for further analysis.

This would require a technological solution that is readily accessible, usable and affordable to the heath facilities.

A way forward

A successful implementation would have the net effect of crowd sourcing important clinical data such as;

1) Patient demographics – age, gender
2) Location
3) Symptoms
4) Diagnosis
5) Drugs prescribed
6) Tests and procedures
7) Amount paid for services where applicable

What could we potentially learn from this data?

We can learn a lot about disease occurrence and spread, occurrence of communicable and non communicable diseases, periodicity of occurrence / outbreaks, geographical dispersion and spread.

The Project – Data Management

Big Data Process Stage	Present Challenge	Proposed Solution
Acquisition and recording	Data is collected using diverse and incompatible systems.	Develop a mobile application that will be accessed on a pay per use approach.
	Data from small facilities is collected manually for keying in the government systems.	Roll out the app to at least 1,000 small clinics, pharmacies, x-ray and lab facilities

The Project – Data Management

Big Data Process Stage	Present Challenge	Proposed Solution
Extraction, cleaning and Annotation	Data collected is incomplete and inconsistent, containing multiple identifiers for patients.	Use patient mobile number or national ID number as the key identifier for all records
Integration, aggregation and Representation	Data collected cannot be correlated and matched to a single individual.	Allow for the sharing of patient data across facilities with patient consent.

The Project – Data Analytics

Big Data Process Stage	Present Challenge	Proposed Solution
Modeling and analysis	Data available is not accurate and timely.	Data synchronization to be in near real time and facilitated by incentives such as per record data refund.
Interpretation	Reports generated are months old and incapable of facilitating timely interventions.	Generate near real time maps and reports.

Work in progress

This project is undertaking the development of a simple, mobile based solution for use in clinic management and data collection to facilitate the collection of data from informal and rural settlements whose data is currently not available.

This data will in turn facilitate analytics and visualization for effective decision making in public health.

Work in progress – Emed Clinic Management System

Thanks