Innovative Open Source Approaches to Automating Clinical Data Management for a Large Multi-centre Cohort Study

Narshion Matai Ngao

Childhood Acute Illness & Nutrition Network,
KEMRI Wellcome Trust Research Programme

KEMRI | Wellcome Trust

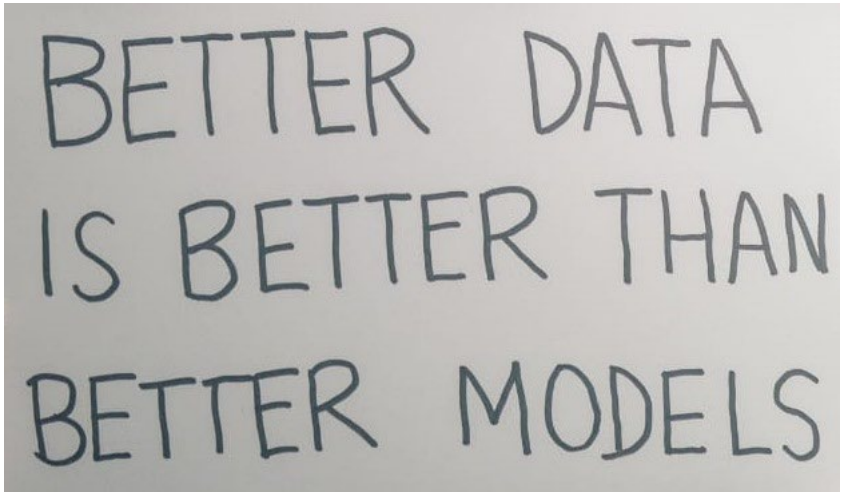
25-11-2022

The Global Health Network Conference 2022, University of Cape Town, South Africa

THE GLOBAL HEALTH NETWORK CONFERENCE SOUTH AFRICA 2022

Data Management in Research Studies

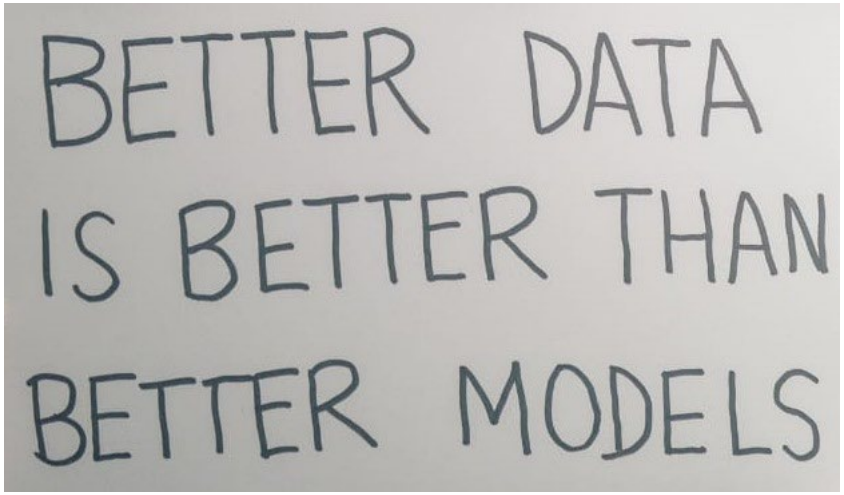
- Do you agree with this?



Source: twitter.com random user post

Data Management in Research Studies

- Do you agree with this?



Source: twitter.com random user post

- **Accurate, Reliable and Statistically sound** data for analysis and decision making

CHAIN Network Data Issues Experience

- A Multi-centre, longitudinal (multi-event) cohort study

- We experienced these data related issues: -
  1. CRF design issues & database laxity – allowing errors to occur. (unforeseen at first)
  2. Inconsistent values entered (wrong dates, abnormal measurements)
  3. Missing data
  4. Programmer issues – formatting, conversion errors, data type mapping, renaming issues
  5. Time to chase all these in a multi-centre environment

Unique Challenges of a Multi-centre Study

Complexity of set up – clinical, social, laboratory data

Infrastructural Environment – different capacity at sites

Data Cleaning turn-around - Query timings

Lack of standards/Standardization – Consistency of approach

Data Reporting – Single point of truth

What we need to see happen

Monitor real time data quality checks.

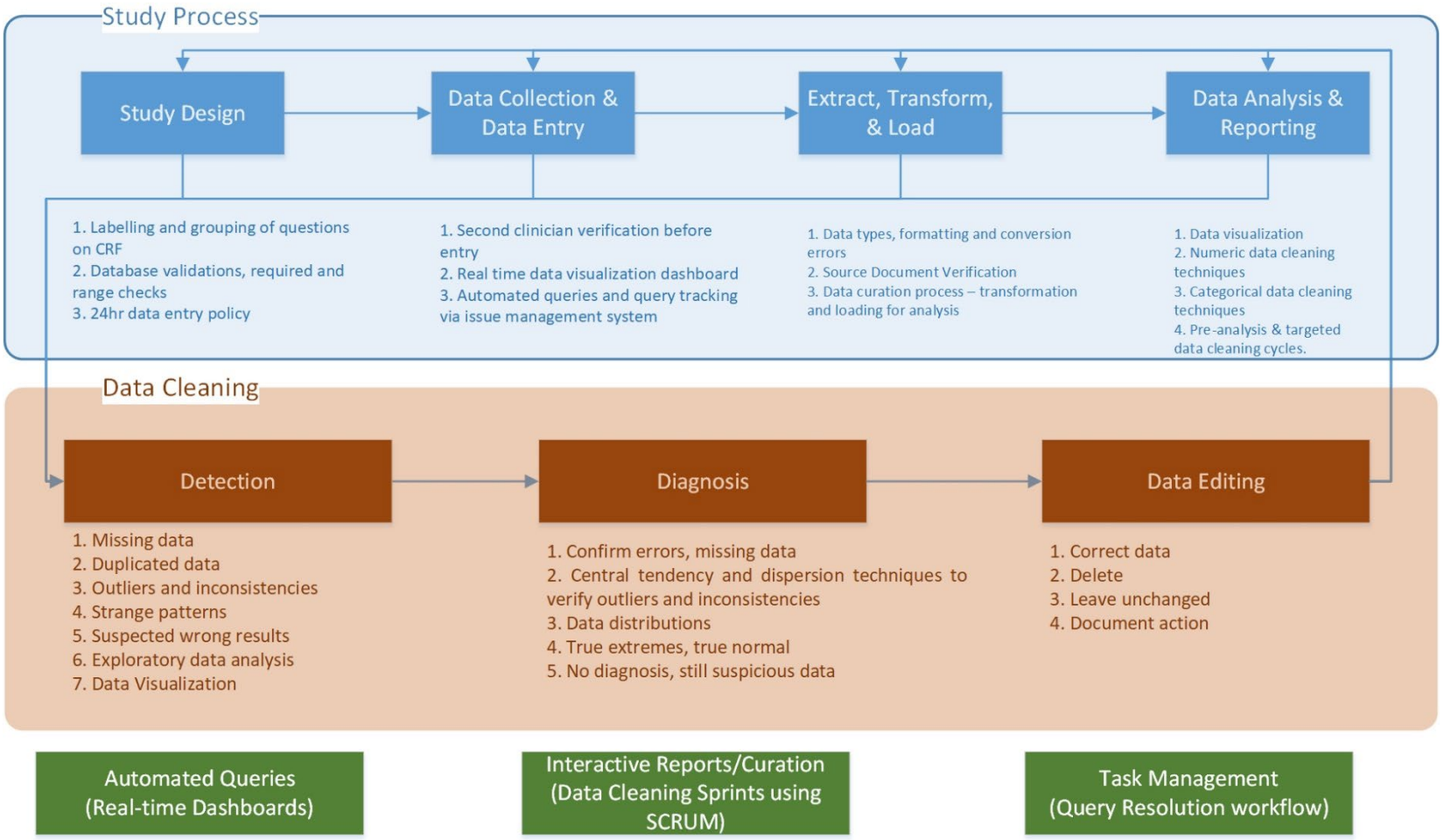
Run data cleaning queries & post them automatically.

Access to Standardized Study Reports

Improve productivity of data team

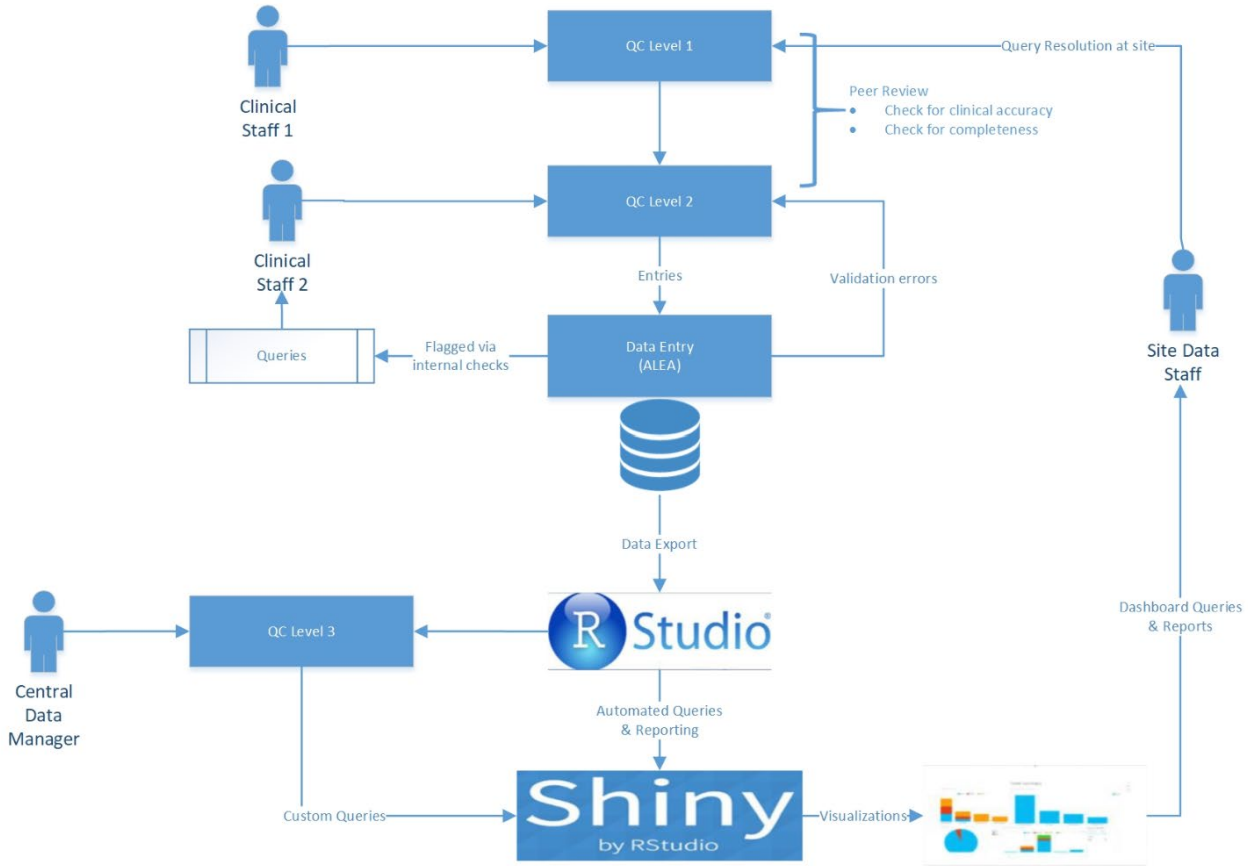
Methods

Solution: Data Management as an Embedded Service



Methods

How We Did It



Our Tools of Trade

- REDCap is a mature, secure web application for building and managing online surveys and databases.
- Kilifi Integrated Database Management System. Primary storage for laboratory meta data and test results.
- Rstudio and Shiny: Great combination of both packages & frameworks to produce a dashboard.

REDCap
Research Electronic Data Capture

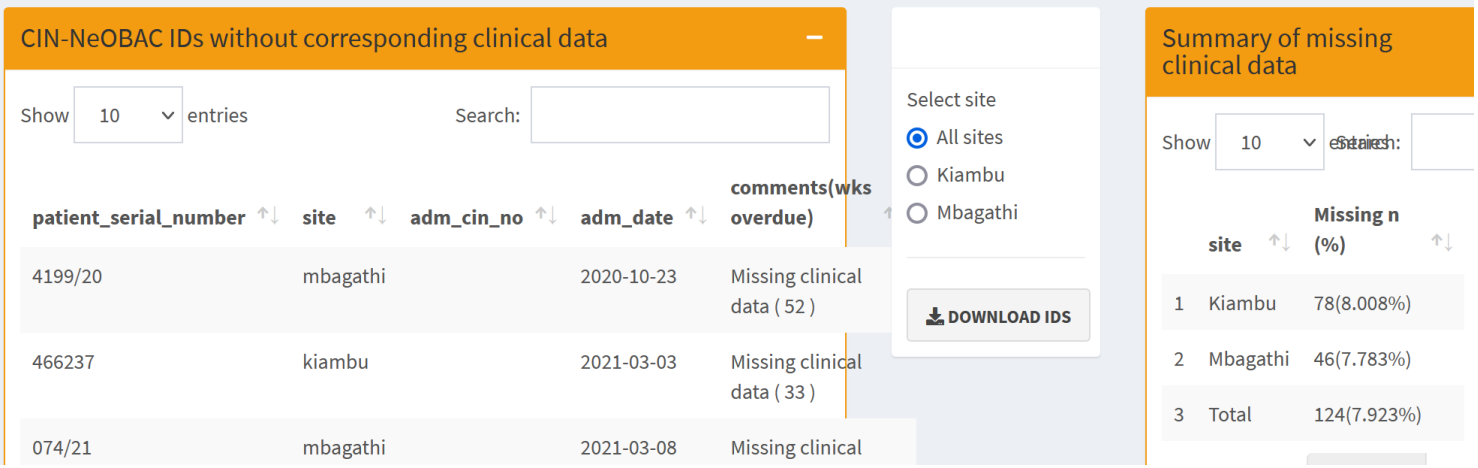
RDBMS (KIDMS)

RStudio

Shiny

Benefits of our approach

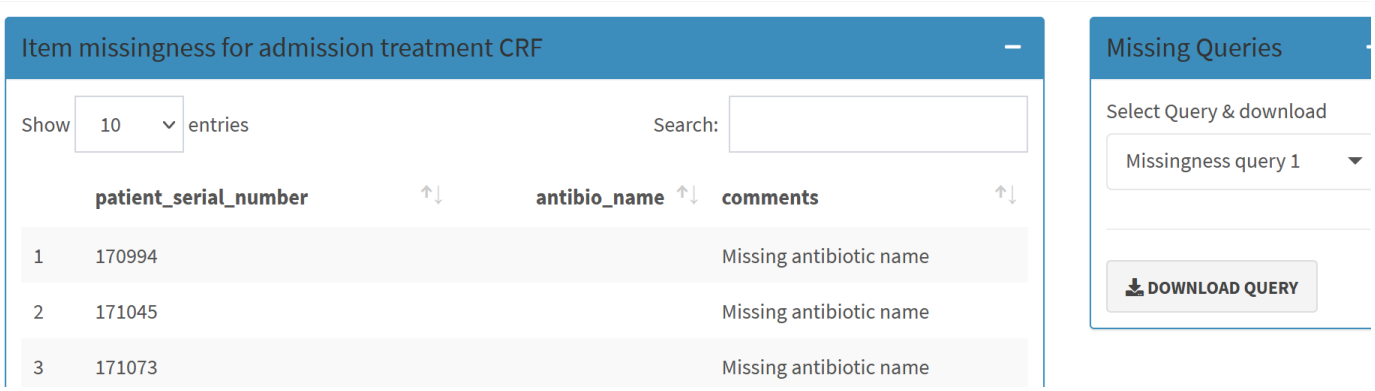
1. A sub-set of risk-based approaches to data management.
2. Real-time check on both item and variable missingness.
3. Up-to-minute capturing of data inconsistency and anomaly detection.
4. Standardized yet interactive study progress reports (accrual curves, lost to follow up and missed visits summaries).
5. Customized utility reports for planning future events such as enrolment targets



THE GLOBAL HEALTH NETWORK CONFERENCE SOUTH AFRICA 2022

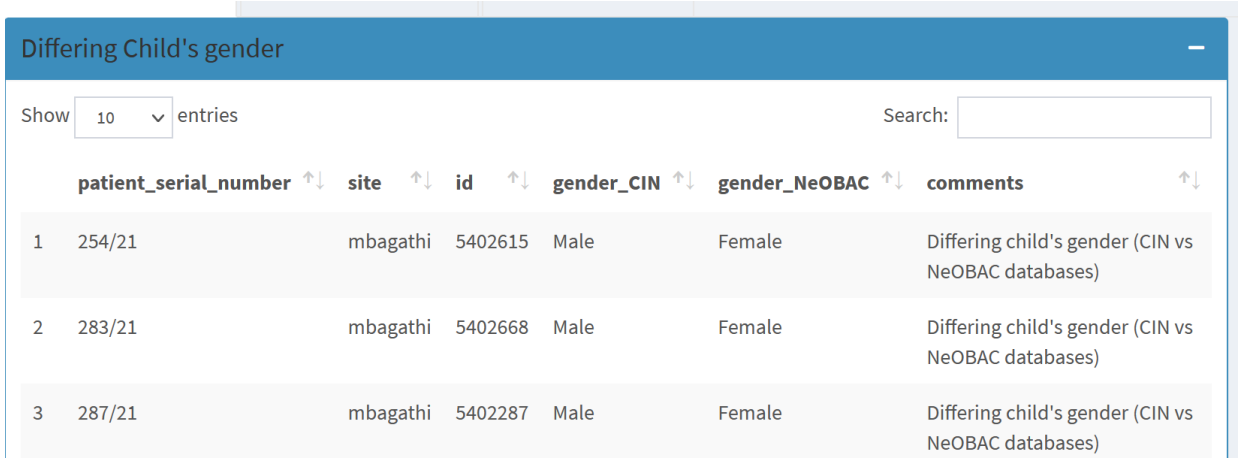
Benefits of our approach

1. A sub-set of risk-based approaches to data management.
2. Real-time check on both item and variable missingness.
3. Up-to-minute capturing of data inconsistency and anomaly detection.
4. Standardized yet interactive study progress reports (accrual curves, lost to follow up and missed visits summaries).
5. Customized utility reports for planning future events such as enrolment targets



Benefits of our approach

1. A sub-set of risk-based approaches to data management.
2. Real-time check on both item and variable missingness.
3. Up-to-minute capturing of data inconsistency and anomaly detection.
4. Standardized yet interactive study progress reports (accrual curves, lost to follow up and missed visits summaries).
5. Customized utility reports for planning future events such as enrolment targets



Differing Child's gender

Show 10 entries Search:

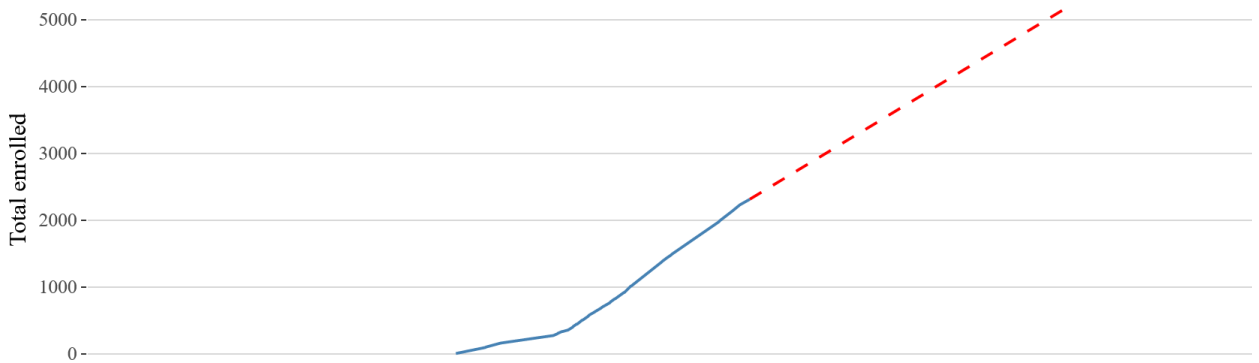
	patient_serial_number	site	id	gender_CIN	gender_NeOBAC	comments
1	254/21	mbagathi	5402615	Male	Female	Differing child's gender (CIN vs NeOBAC databases)
2	283/21	mbagathi	5402668	Male	Female	Differing child's gender (CIN vs NeOBAC databases)
3	287/21	mbagathi	5402287	Male	Female	Differing child's gender (CIN vs NeOBAC databases)

Benefits of our approach

1. A sub-set of risk-based approaches to data management.
2. Real-time check on both item and variable missingness.
3. Up-to-minute capturing of data inconsistency and anomaly detection.
4. Standardized yet interactive study progress reports (accrual curves, lost to follow up and missed visits summaries).
5. Customized utility reports for planning future events such as enrolment targets

Benefits of our approach

1. A sub-set of risk-based approaches to data management.
2. Real-time check on both item and variable missingness.
3. Up-to-minute capturing of data inconsistency and anomaly detection.
4. Standardized yet interactive study progress reports (accrual curves, lost to follow up and missed visits summaries).
5. Customized utility reports for planning future events such as enrolment targets



Results and Conclusions

Seamless data extraction and processing with no human intervention.

Run automated cleaning scripts and monitor data quality checks in real time.

Up-to-minute access to standardized study reports by sites' PI and study sta.

Real time study progress reports helped planning of recruitment and other activities.

Improved productivity of the data team and reduced over-reliance on the central team

Recommendations

For a Multi-centre Study: -

1. Standardizing data management practice across all participating sites
   - Ensure a uniform data collection and management environment for all the sites.
2. Database for data collection should be hosted online (via internet) so that it is the same or if using offline data capture, then data should be synced to the same online server as soon as possible.

3. Real-time (or almost) data monitoring and query resolution dashboards be implemented and hosted online.
   - Cut back data managers time spent on repetitive tasks & standard reports.
4. Participating sites to take up consumption of the centralized dashboards whenever they're availed.

Acknowledgements

BILL & MELINDA GATES foundation

UNIVERSITY OF OXFORD

KEMRI | Wellcome Trust

Childhood Acute Illness & Nutrition Network

DEPARTMENT OF GLOBAL HEALTH
UNIVERSITY of WASHINGTON

THE AGA KHAN UNIVERSITY

icddr,b

KENYA MEDICAL RESEARCH INSTITUTE
KEMRI

MULAGO HOSPITAL

SickKids
Centre for Global Child Health

Mbagathi Hospital
NAIROBI CITY COUNTY HEALTH SERVICES

Moyo Nutritional Rehabilitation & Research Unit
College of Medicine, Blantyre, Malawi

Migori County Hospital
MIGORI COUNTY HEALTH SERVICES

UNIVERSITY OF MALAWI

CIVIL HOSPITAL KARACHI

The Global Health Network Conference 2022, University of Cape Town, South Africa

Special thanks to the patients, staff and leadership at our clinical sites.