MEIRU
Malawi Epidemiology and Intervention Research Unit

Mind the Gap – Modernising existing data documentation for a long-standing research institution in Malawi

Case study on application of DDI codebook to longitudinal demographic, epidemiological and bio-medical data

Chifundo Kanjala, Jacky Saul, Estelle McLean, Keith Branson, David Beckles, Mia Crampin

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Background

Malawi

- Located in in Southern Africa
- Pop size approximately 17 million
- Under 1000 doctors serving the entire country (unofficial sources)
- High HIV prevalence country at about 10.6%
- Increasing burden of NCDs
- Among least resourced countries

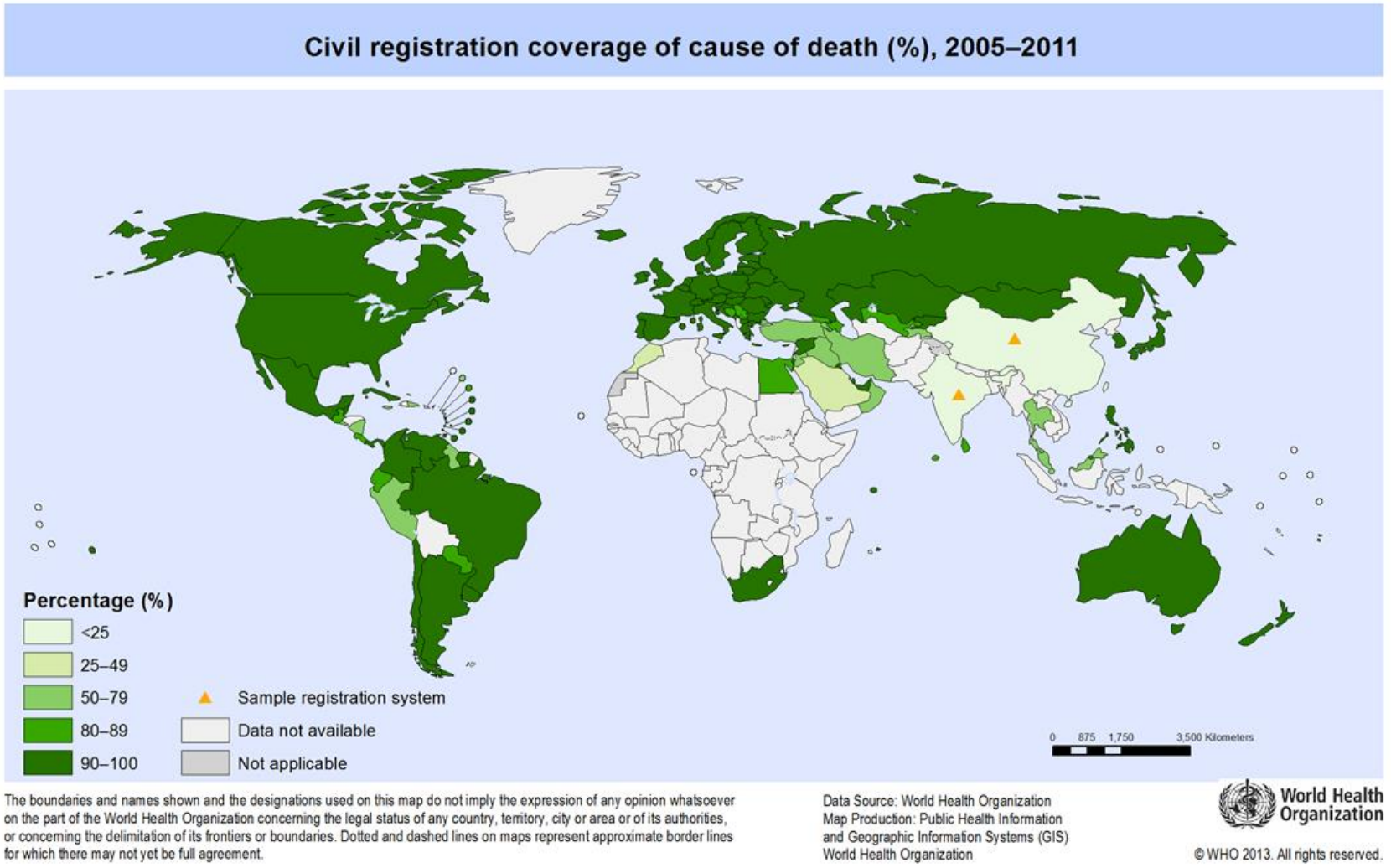
LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Malawi Epidemiology and Intervention Research Unit

- Focus on population health research
- Partnership between Malawi College of Medicine, London School of Hygiene and Tropical Medicine (LSHTM) and Malawi Ministry of Health
- 1979 – present, rural northern Malawi
  - District level health data ~300,000 individuals,
  - Data linked across time and studies
  - Socio-demographic and health surveillance data~ 40,000 individuals
  - Historically focused on infectious diseases
- 2013 – present, urban site Lilongwe
  - Non-communicable disease data

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Why bother?



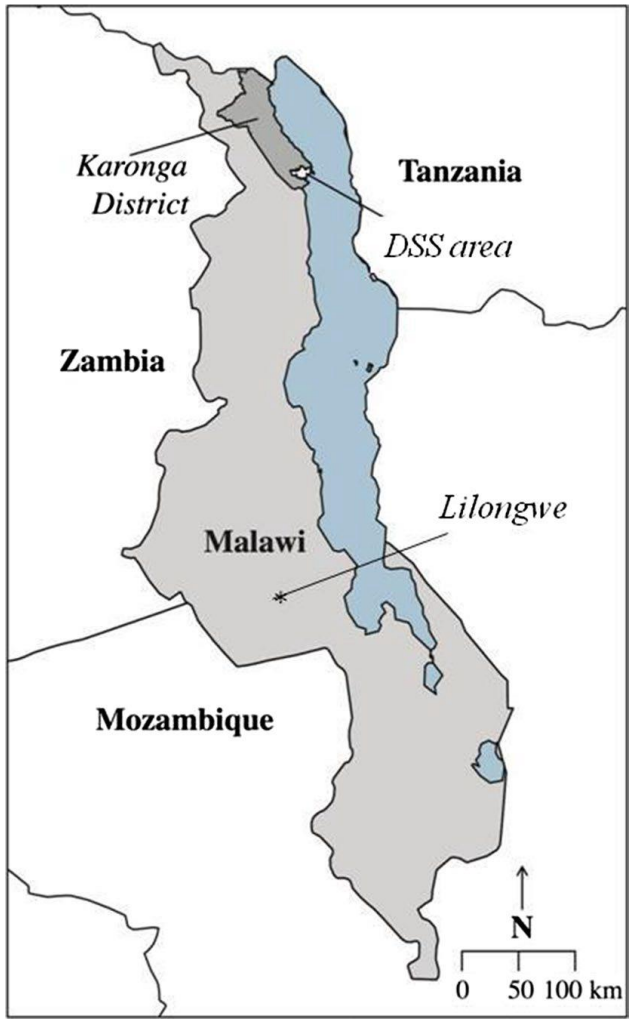
Issues

- Malawi is among least developed countries
- Lack of burden of disease data
- Population-based data production is difficult
- Lack of models for preparing legacy data for sharing

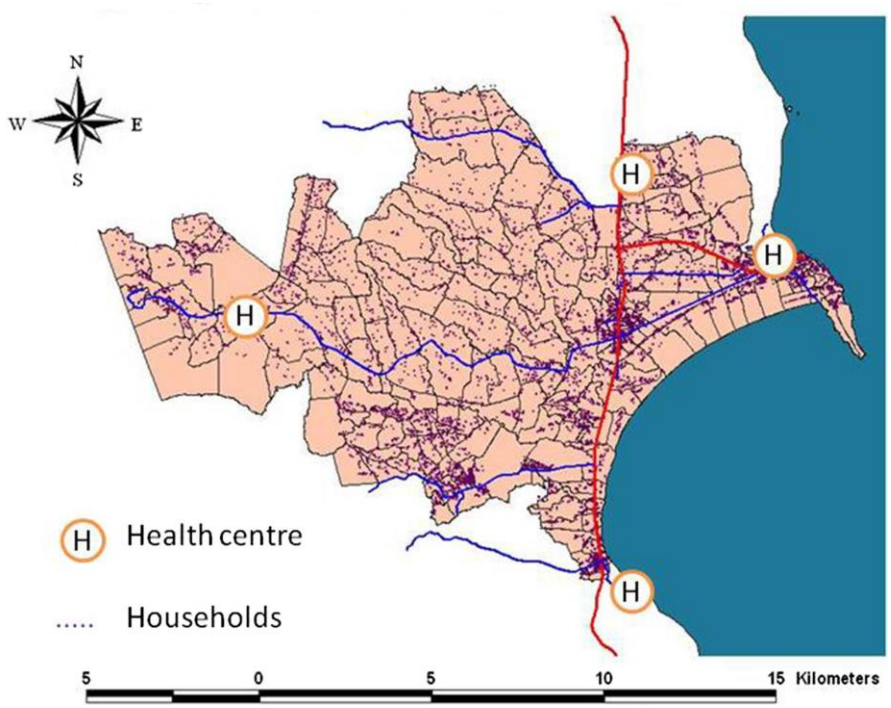
Addressing issues

- Our research programme contributes to fill data gap
- Write up our experiences as a case study

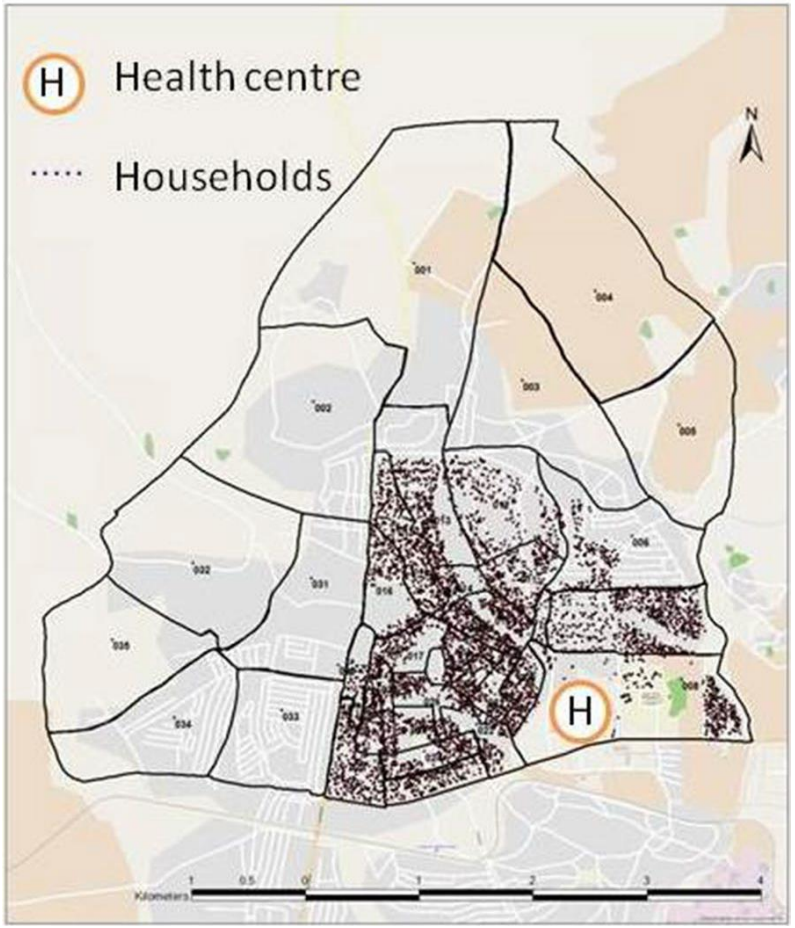
Malawi

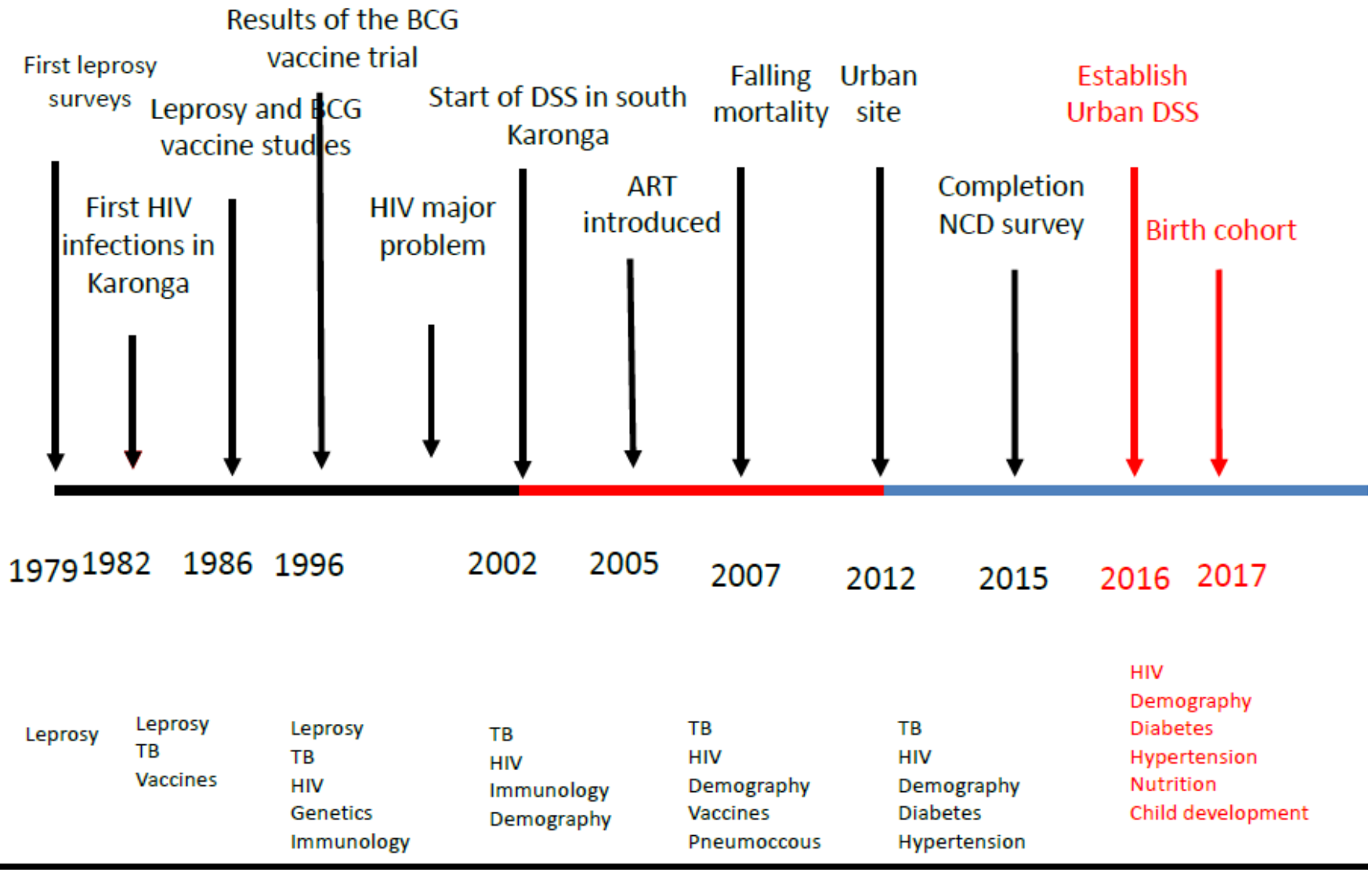


Karonga rural site



Lilongwe urban site




First leprosy surveys
Leprosy and BCG vaccine studies
Results of the BCG vaccine trial
First HIV infections in Karonga
HIV major problem
Start of DSS in south Karonga
ART introduced
Falling mortality
Urban site
Completion NCD survey
Establish Urban DSS
Birth cohort
1979
1982
1986
1996
2002
2005
2007
2012
2015
2016
2017
Leprosy
Leprosy
TB
Vaccines
Leprosy
TB
HIV
Genetics
Immunology
TB
HIV
Immunology
Demography
TB
HIV
Demography
Vaccines
Pneumoccous
TB
HIV
Demography
Diabetes
Hypertension
HIV
Demography
Diabetes
Hypertension
Nutrition
Child development

Data Status

- Main datasets in several MS Access databases
- Currently migrating to Postgres
- 500+ tables, 25,000+ variables
- Supported by validation databases, coding manuals, data manuals, entity relationship models, ODK data dictionaries
- Also composite/ analytical datasets

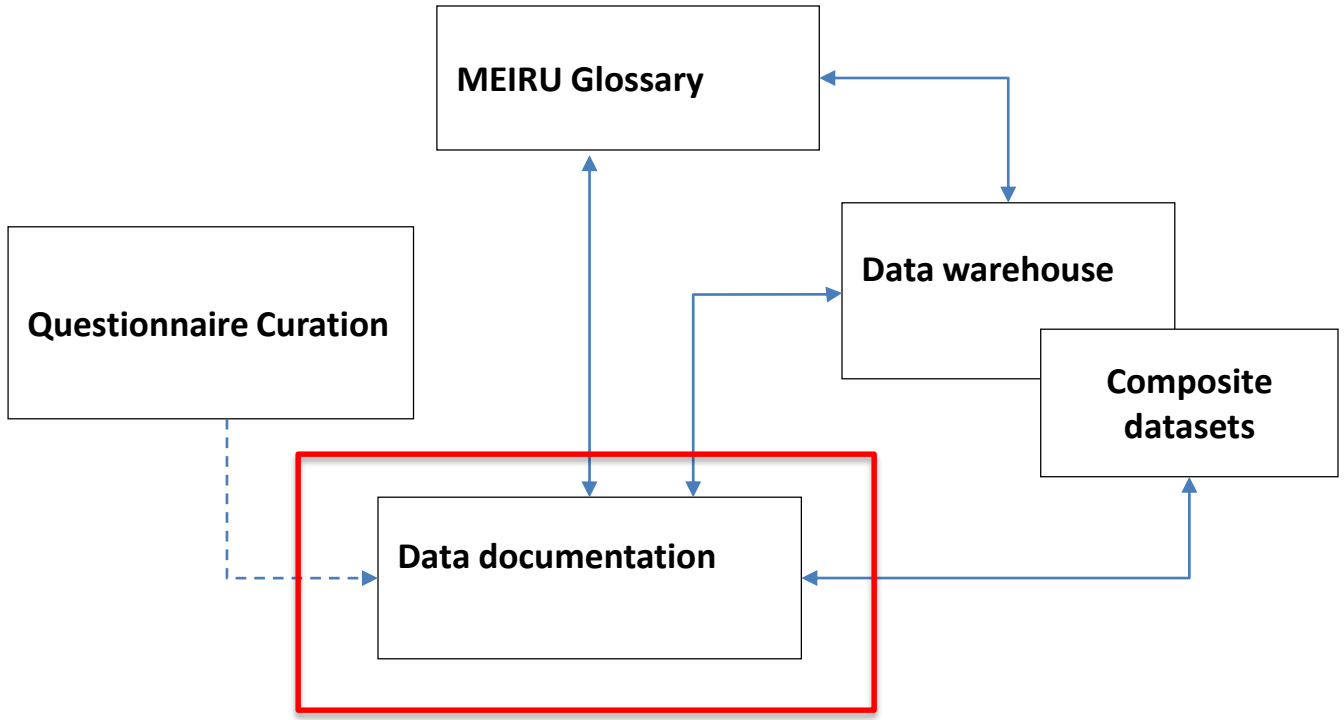
LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Objective

- Maximising usefulness of MEIRU data

Targeted deliverables

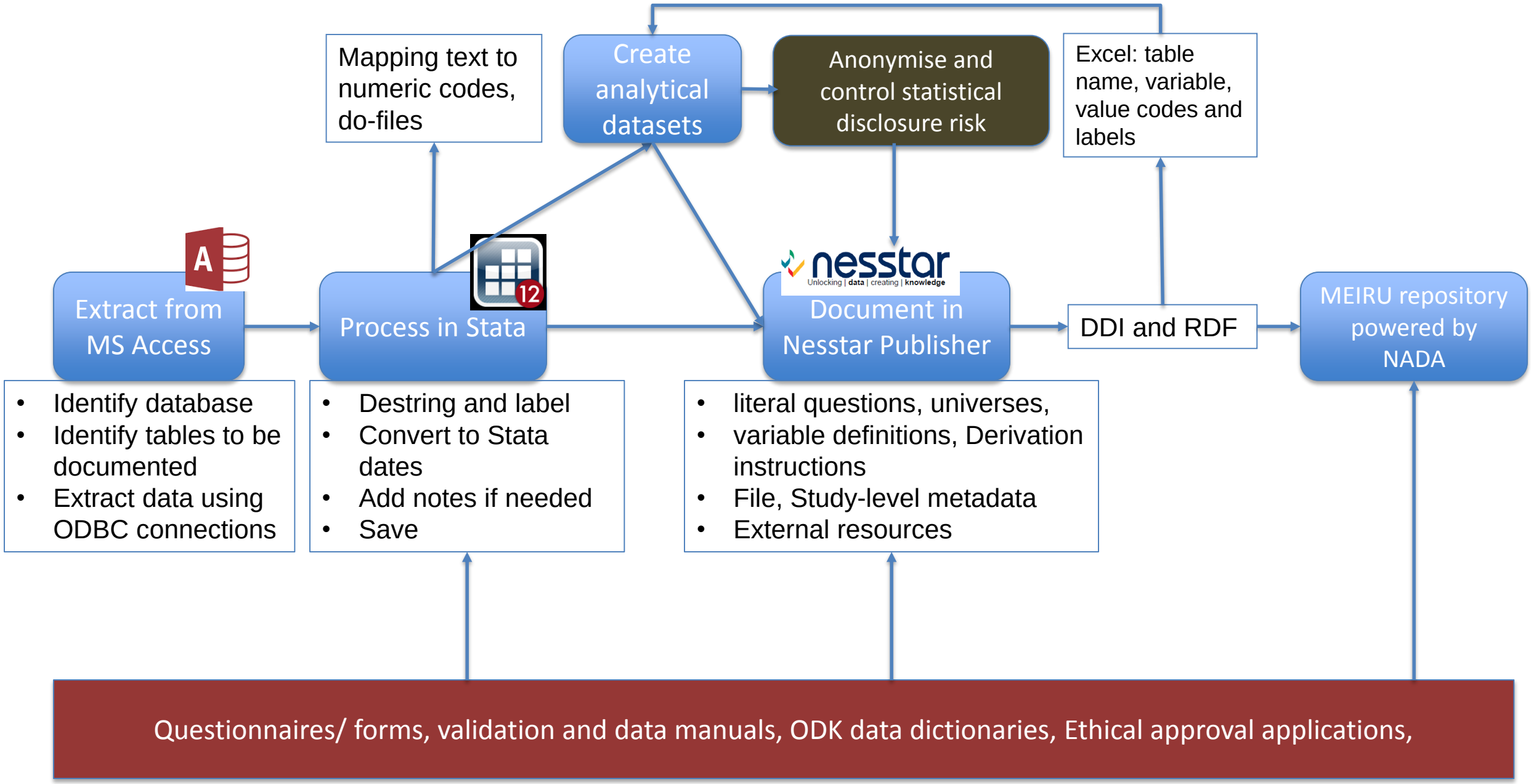
- New data structure for easy access
- Documentation using metadata standards
- Development of data sharing policies
- Efficient methods for data capture
- Lodging subset of data in public repository
- Incorporation of urban site data



LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Where are we now?

Data documentation workflow



- Trained metadata officers in Stata, Nesstar and NADA
- Documented cleaned data in major databases
- Setup World Bank developed NADA catalogue
- Initial discussion on documentation of analytical datasets
- Identified LSHTM data compass for data sharing

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Documenting analytical datasets

- 8 categories of variables related to analytical datasets identified
- Distinction between documentation for internal use and for shared data
- We are using the data file description and derivation and imputation fields, Nesstar Publisher
- Data file description used to define categories of variables
- Mainly for human consumption not for guiding software apps

	Code	Description
1	Sx	From Source but not used
2	Sd	From source, used but then dropped
3	S	From source, kept
4	Dpx	Previously derived but not used
5	Dpd	Previously derived variable, used but then dropped
6	Dp	Previously derived variable, kept
7	Dd	Derived variable, used but then dropped
8	D	Derived variable, kept

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Achievements

- Consistent pattern in MEIRU manuals thus easy to train officers
- High quality data stored securely
- Data management team with vast experience and high competence
- Accessible gateway for data, study documents and study outputs
- Availability of tools to use – Nesstar and NADA
- MEIRU data will be more accessible internally and externally

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Challenges

- Study level metadata not easy to extract
- Bulk of data collected on paper
  - We can't leverage advantages afforded by electronic data capture
- Time required to create/ edit metadata for complex studies
- Ongoing work – updates needed for new data
- Robust versioning to keep track of do-files, xml, extracted data and changes to manuals

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Where to next?

- Documentation of analytical datasets
- Adding all documented data to MEIRU data portal
- Updating of data sharing policies to guide data access requests
- Anonymisation and statistical disclosure risk assessment and control
- Depositing data subset with LSHTM data compass

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

Acknowledgements

- MEIRU study participants
- MEIRU data office colleagues and field interviewers
- MEIRU Metadata officers
- Colleagues in DDI community
- Wellcome Trust
- ALPHA network – London School of Hygiene and Tropical Medicine
- INDEPTH network
- College of Medicine, University of Malawi
- Malawi, Ministry of Health

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE