Outline

- Data Quality Indicators
  - Workshop 11 – 13 May 2010 Accra
- Metadata Technology Review of iShare
- The Way Forward
Data Quality Metrics for Minimum Micro Dataset

- **Attribute Domain Indicators**
  - Measure whether all dataset variables are present and their values valid
    - Key Indicators (proportion of:)
      - Individuals with mother identity specified
      - Deaths with cause coded in ICD-10
      - Births with precision at day level

- **Relational Integrity Indicators**
  - Verify that all references between minimum dataset components are consistent
    - Key Indicators (proportion of:)
      - Individuals with at least one residency episode
      - Deaths linked to an individual
      - Births linked to an individual
      - Births linked to a pregnancy that is linked to an individual
      - Individuals with similarity measure >2
Data Quality Metrics for Minimum Micro Dataset

- **Historical Data Indicators**
  - Data Currency
    - Key Indicators
      - Proportion of current residents observed during the last complete surveillance round
  - Observation Granularity
    - Key Indicators
      - Proportion of visits gaps (duration between subsequent visits to same homestead) falling within 10% deviation of the surveillance round duration
  - Event Histories
    - Key Indicators
      - 1 - Proportion of births to the same woman spaced at less than 196 days (28 weeks)
      - Proportion of births that are to women between the ages of 12 and 49yrs of age
Data Quality Metrics for Minimum Micro Dataset

- State Transition Rules
  - Terminator State Constraints
    - Key Indicators
      - Proportion of Individuals with valid states at first transition
  - State Transition Constraints
    - Key Indicators
      - Proportion of individuals with valid residency state transitions
  - State Duration Constraints
    - Key Indicators
      - Proportion of individuals with residency state durations greater than zero
  - Action pre-conditions
    - Key Indicators
      - Proportion of residencies started with a birth where the mother is resident at the time of birth

- Attribute Dependency Rules
  - Key Indicators
    - Proportion of births linked to mother via pregnancy that is consistent with mother identity on child’s record, and converse
    - Demographic balance equation: Correspondence between calculated resident population at end of year with measured resident population at start of subsequent year

INDEPTH Network
Metadata Technology

ISHARE REVIEW
iSHARE

- Significant progress and contributions made towards improving access to harmonized datasets.
- Identified and addressed data quality issues.
- Coordinated information exchange and collaboration with participating centres.
- iSHARE web platform is functional and demonstrated what a fully developed site could be capable of.
- iSHARE team has gained considerable expertise
- Laid the foundation of a data harmonization and sharing framework
- Cultivated the right ideas for data sharing
- Demonstrated that bringing together data from the multiple sources is possible.
- Shown that such a task is not a trivial one.
Challenges

- Meeting the needs of all stakeholders, from centre level to external research community and sponsors
- Improving overall data quality and documentation
- Further examining harmonization and comparability issues
- Providing a flexible platform that can be used at both surveillance centres and centrally, is adapted to local capacity, and can operate in a federated environment
- Adopting data access and sharing policies that meet the needs of all data providers
- Ensuring the protection of confidential respondent data through sound statistical data disclosure practices
- Making the project sustainable by strengthening internal capacity and expertise
- Extending the vision beyond data management by providing a platform that fosters collaborative research and knowledge sharing
Recommendations

- Adopt Data Documentation Initiative (DDI) specification as metadata format and an open text format for the exchange, preservation and dissemination of data.
- Adoption and integration of loosely coupled data/metadata management tools for use at centres and Network level.
- Deployment of federated web based catalogues to support the discovery of centre and Network level data, deliver comprehensive data documentation, and manage access to underlying datasets.
- Leverage DDI metadata to maximize automation of underlying processes, improve timeliness, and increase overall data quality.
- Maintenance of reference metadata at Network level to foster and ensure data consistency and quality.
- Ensure the availability of an easy to install and maintain hardware/software solution so that relevant tools can be deployed at all centres.
THE WAY FORWARD
INDEPTH Data System Initiatives

- Establish a detailed database of member centre capacity
  - INDEPTH Member Survey
- Promote the adoption of core data quality metrics
- Support initiatives to develop common and next generation data management systems
- Support and expand the iSHARE initiative
INDEPTH Strategic Award Proposal

- 2009 Proposal to Wellcome Trust not successful
- Wellcome Trust provided funding for proposal development and re-submission in 2010
- New proposal being developed (pre-proposal submitted in August)
  - Strengthening and extend iSHARE based on review recommendations
  - Build data management capacity by introducing a data management track in the INDEPTH MSc Leadership Programme.
Centre-in-a-Box

Data Managers

External Users

Local Users

Centre-in-a-box

OpenHDS Server

Data/Metadata Desktop

Study Catalog Server

... SQL Database Server

Data Analysis Desktop

Web Server

Admin Desktop

Secure Storage

Remote Admin

INDEPTHTH Network