Harmonising data collection, management and sharing – the INDEPTH Experience

INDEPTH Scientific Conference
11-13th Nov 2015, Addis Ababa, Ethiopia

Kobus Herbst\textsuperscript{1,2}, Nicolas Maire\textsuperscript{4}, Sanjay Juvekar\textsuperscript{1,3}, Osman Sankoh\textsuperscript{1}

1. The INDEPTH Network, Ghana
2. The Africa Centre for Population Health, UKZN, South Africa
3. KEM Research Centre, Vadu, India
4. Swiss Tropical and Public Health Institute
Introduction

IDMP

INDEPTH Data Management Programme
OpenHDS

Implementation

- INDEPTH Centres
  - Cross River
  - Ifakara Health Institute
  - Manhica
  - Nanoro

- INDEPTH associate members
  - Rusinga Island

- Other disease surveillance sites
  - Majete
  - Bioko Island

- Implementation in progress (Ethiopian sites):
  - Arba Minch, Butajira, Dabat, Gilgel Gibe, Kersa, Kilite Awlælo
Major Developments

• Household level visits catering for more than one household per dwelling
• Internationalisation, including support for multiple languages, scripts and calendars
• Extensibility of core data collection for site specific questionnaires
• Data Migration tool chain
• Collaborative infrastructure to enable mutual support and learning
  – Common code repository
  – Issue/bug tracker
  – Mailing list
Lessons Learned

• Complexities of Data Migration
  – Non-standard systems
  – Data cleaning
  – Support for non-core variables

• Ongoing requirement for software maintenance
  – OpenHDS build on a number of open source components with their own release cycle requiring re-integration of components

• Need to decrease the dependency on the Swiss TPH team by further strengthening capacity and exchange in the INDEPTH OpenHDS community
Challenges

• Maintaining a technical and implementation support capacity
  – Central capability in the short to medium term
  – Strengthening INDEPTH centres’ technical capacity in the longer term to decrease dependence on central support

• Need to work hard at maintaining standard database structure and common code-base because this is the real strength of OpenHDS
INDEPTH Data Repository
Highlights

• INDEPTH Cause Specific Mortality - Oct 2014
  – 22 sites, 111,000 deaths

• July 2015 Core Data Update
  – 27 sites participated, 20 datasets shared

• Recognition by PLoS as a repository to host publication data sets

• Publication of iSHARE procedures and design in *J Empir Res Hum Res Ethics*
INDEPTH Repository: Shared Individual Level Data

Cause of Death
- 111,910 Deaths
- 98,429 Verbal Autopsies
- 22 Sites

Person Years

- July 2013: 6 Sites
- July 2014: 13 Sites
- July 2015: 20 Sites
- In Progress: 25 Sites

Millions
Who are downloading?
Future Plans

• IDMP workshop in April 2016
  – New Centre’s joining iSHARE
  – Update of existing core datasets
  – Update of cause-specific mortality datasets
  – Expanding core dataset with education variable

• Final workshop in 2017
Highlights

• July 2015 Update
• INDEPTHStats App for Android and iOS
Acknowledgements

Wellcome Trust
Strategic Award

The William and Flora Hewlett Foundation
Bill & Melinda Gates Foundation
Sida
IDRC CRDI

iSHARE 2
an INDEPTH project
Disclosure Risk Reduction

- Event history microdataset with exact date of birth and event dates
  - Anonymised Id A
  - Public Access

- Cause of death microdata with year of death and age group
  - Anonymised Id B
  - Public Access

- Identity Map
  - Anonymised Id A
  - Anonymised Id B
  - Restricted Access

iSHARE 2
an INDEPTH project
Centre-in-a-Box

Database Server

Data-Manager Desktop

ETL
Publisher
R

Web Server, Data Repository (NADA), File Manager

System Administration

Remote Support & Help Desk

iSHARE 2
an INDEPTH project
<table>
<thead>
<tr>
<th>Attribute</th>
<th>Variable name</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Record number</td>
<td>RecNr</td>
<td>A sequential number uniquely identifying each record in the data file</td>
</tr>
<tr>
<td>Centre identifier</td>
<td>CentreId</td>
<td>An identifier issued by INDEPTH to each member centre of the format CCCSS, where CCC is a sequential centre identifier and SS is a sequential</td>
</tr>
<tr>
<td></td>
<td></td>
<td>identifier of the site within the centre in the case of multiple site centres</td>
</tr>
<tr>
<td>Individual identifier</td>
<td>IndividualId</td>
<td>A number uniquely identifying all the records belonging to a specific individual in the data file. For data anonymization purposes, this number</td>
</tr>
<tr>
<td></td>
<td></td>
<td>should not be the same as the identifier used by a contributing centre to identify the individual, but the contributing centre should retain a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>mapping from this identifier to their identifier</td>
</tr>
<tr>
<td>Country identifier</td>
<td>CountryId</td>
<td>ISO 3166-1 numeric code of the country in which the surveillance site is situated</td>
</tr>
<tr>
<td>Location identifier</td>
<td>LocationId</td>
<td>Unique identifier associated with a residential unit within the site and is the location where the individual was or became resident when the</td>
</tr>
<tr>
<td></td>
<td></td>
<td>event occurred. For data anonymization purposes, this identifier should not be the same as the identifier used internally by the contributing</td>
</tr>
<tr>
<td></td>
<td></td>
<td>centre, but the contributing centre should retain a mapping of this identifier to their internal location identifier</td>
</tr>
<tr>
<td>Date of birth</td>
<td>DoB</td>
<td>The date of birth of the individual</td>
</tr>
<tr>
<td>Event</td>
<td>EventCode</td>
<td>A code identifying the type of event that has occurred (Table 2)</td>
</tr>
<tr>
<td>Event date</td>
<td>EventDate</td>
<td>The date on which the event occurred</td>
</tr>
<tr>
<td>Observation date</td>
<td>ObservationDate</td>
<td>Date on which the event was observed (recorded), also known as surveillance visit date</td>
</tr>
<tr>
<td>Event count</td>
<td>EventCount</td>
<td>The total number of events associated with this individual in this data set</td>
</tr>
<tr>
<td>Event number</td>
<td>EventNr</td>
<td>A number increasing from 1 to EventCount for each event record in order of event occurrence</td>
</tr>
</tbody>
</table>