Legacy data migration: Introduction

nicolas.maire@unibas.ch
Illustrative example: Ifakara Health Institute HDSSs*

Ifakara Health Institute has 2 sites

Ifakara

  Operational since September 1996
  Urban DSS with approx 16,000 households followed via 2 rounds per year
  Rural DSS with approx 58,000 households followed via 2 rounds per year

Rufiji

  Operational since Nov 1998
  DSS with approx 41,000 households followed via 2 rounds per year

*Slides: Tumaini Kilimba, IHI
Migration Process

Transform HRS2 data table from DBF to CSV
1 hours

Import data into MySQL “staging” tables
2 hours

Transfer data from staging tables to OpenHDS tables via OpenHDS web services.

Up to 12 weeks (considering the data cleaning iteration in next slide)

Rejection of any data not conforming with constraints placed on entity attributes.
Rejected data is exported into Excel with descriptions of what is wrong with it.

This is sent back to data managers who fix the issues by going through the relevant paper stored version and sending a fieldworker to verify/correct the data at source.

The amendments are sent back to us and once again tested for consistency.

Clean data is let through, inconsistent data is rejected.

Rejected data is again exported to Excel and sent back to data managers.

Continuous iterative process until no more can be done. Hence a useful by-product of HRS2-OpenHDS migration is cleaner legacy data.
**MIGRATION PROCESS**

1. **STEP 1: TRANSFORM**
   DBF TABLES TO CSV

2. **STEP 2: IMPORT**
   INTO MySQL
   STAGING TABLES

3. **STEP 3: TRANSFER**
   DATA FROM
   STAGING TABLES
   TO OpenHDS
   USING MIRTH

   **OPENHDS REJECTS DATA?**
   - **NO**
     - CLEAN DATA IN OpenHDS DB
   - **YES**
     - **REJECTED DATA EXPORTED TO EXCEL,**
       SENT TO DATA MANAGERS FOR CLEANING

   **REPEAT STEP 2**
The “data-migration->data-cleaning->data-migration” cycle is painstakingly slow and laborious (but worth it!).

Some data is beyond recovery (the individuals/households concerned cannot be traced, and an informed correction becomes impossible)
Available tools and experience

Data migration tools from HRS2 to OpenHDS
Virtualized servers with all necessary pre-requisites installed
Documentation: a section in the OpenHDS manual
A demo/tutorial: https://github.com/SwissTPH/openhds-from-hrs2

Other data systems will need a preprocessing step
Data migration infrastructure: your server instance

**Server configuration**

Virtual server image (VMWare/Virtualbox)*

Ubuntu 16.04
MySQL 5.7
Tomcat 8
MirthConnect

For training purposes only!

*Thank you Brendan Gilbert, Africa Centre for Population Health!
Additional prerequisites

Server

R
For data migration

Python
For monitoring tools and simulation

Pentaho
For iShare integration

Client

Java
For Mirth Connect Administration

SSH Client
e.g. MobaXterm
(http://mobaxterm.mobatek.net/)

Staging DB

OpenHDS

Sync Data (Mirth)

MySQL

MySQL

Server
Data connections, including data migration

- ODK Aggregate
- Sync Data (Mirth)
- OpenHDS
- Baseline Channels
- Update Channels
- Migration Channel
- Staging DB
- HRS2
Goals for today

Have a running server instance for data migration
Make sure this is network accessible (for tablet connectivity)

Either use your CiB to host your instance (VMWare)
Or use the workshop server (Virtualbox)

Images:
http://tinyurl.com/openhds2016

WLAN:
IDMP_O
openhdsdubai
Create new vm choose custom, select VM version 7, configure suitable cpu, mem and nic settings.

When asked select disk choose “use existing virtual disk”. Browse to folder where converted vdisk is located and finish installation.

Power on VM and logon

Edit interfaces file e.g. sudo nano /etc/network/interfaces and change enp0s(x) to ens32, e.g.

Add # The primary network interface
auto ens32
iface ens32 inet dhcp

Reboot vm

Configure network interface as desired i.e. dhcp or static.