

Annex 2: Core minimum micro-data set specifications

Introduction

Any demographic surveillance site requires a standard format for core individual surveillance data in order to calculate individual exposure and basic demographic rates. In the case of INDEPTH Network this format will allow member centres to contribute their surveillance data in a standard way for Network analytical purposes and will facilitate pooling of Network data for multi-site analysis. This format should also allow the calculation of data quality metrics in a standard fashion.

Based on the experience of the INDEPTH Migration, Urbanisation and Health Working group a format based on an event history analysis approach is used as the basis of this specification. This is a flexible format that lends itself to the calculation of basic demographic rates, but is also suitable for event history analysis. This core dataset can be easily expanded through the addition of event attributes or other status events (e.g. educational attainment, employment status or anthropometry measurements). The surveillance data associated with a particular individual over the course of his/her exposure to demographic surveillance are represented by a series of event records. The first event for any individual will either be enumeration, birth or in-migration followed by a variable number of event records for each observed event associated with that individual, such as out-migration or death.

Events

Table 2 defines the different possible events in this specification. Two categories of events are defined, namely, events that changes the residency status of the individual and incidental events (or observations) that do not. Each event has a set of standard attributes (Table 1) that are common to all events, followed by a variable number of attributes specific to the event.

Table 1 : Common event attributes

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

Event	Code	Definition	Attributes	Attribute Description
Birth	BTH	The birth of an individual to a resident female.	MotherId DeliveryEventId	The IndividualId of the mother The RecNr of the delivery event associated with this birth
Enumeration	ENU	Starting event for all individuals present at the baseline census of the surveillance area. It is the date on which the individual was first observed to be present in the surveillance area during the baseline census.		
In-migration	IMG	The event of migrating into the surveillance area	Origin	Classification scheme to be developed by MADIMAH
Out-migration	OMG	The event of migrating out of the surveillance area	Destination	Classification scheme to be developed by MADIMAH
Location exit	EXT	The event of leaving a residential location within the surveillance area to take up residence in another residential location within the surveillance area	Destination	The LocationId of the location within the surveillance area to which the individual relocated

¹ For data production use only

Location entry	ENT	The event of taking up residence in a residential location within the surveillance area following a location exit event. Note that location exit and entry are actually two parts of the same action of changing residential location and as such happens on the same event date	Origin	The LocationId of the residential location from which the individual moved.
Death	DTH	The death of the individual under surveillance. The date of death is the event date.	Cause1 Cause2 Cause3 Likelihood1 Likelihood2 Likelihood3	Up to three causes of death coded using the WHO list of verbal autopsy death causes. Likelihood values associated with each possible cause of death.
Delivery	DLV	The event of a pregnancy end after 28 weeks of gestation, which may or may not result in the birth of one or more individuals (represented in this dataset by a BTH event linked to this delivery event)	LBCnt SBCnt Parity	Live birth count Stillbirth count The number of live births to this women prior to this delivery
Observation end	OBE	An event inserted when a dataset is right censored at an arbitrary date and this individual remained under surveillance beyond this date. The right censor date is the date of this event		
Last Observation	OBL	An event indicating the last point in time on which this individual was observed to be present and under surveillance. Event date equals observation date in this instance. Normally there should be no individuals with this event as their last event if the right censoring date is prior to the start of the last complete census round.		
Observation	OBS	Used to record characteristics of individuals under surveillance valid at the time of the observation. Could be used to record aspects such as educational attainment, employment status or anthropometry measures. Specific examples of this event is not part of the minimum core individual dataset, but is specified to allow for site or working group needs.		