



SOMAARTH DDESS

OPERATIONAL MANUAL

The INCLEN Trust International F1/5, 2nd Floor, Okhla Industrial Area Phase-I, New Delhi- 110020

Page 1 of 49

Contents

Purpose of the Operational Manual	3
Preface	4
Chapter 1: Introduction	5
Chapter 2: Initial Work-up while establishing the DDESS	8
Chapter 3: DDESS: Key Definitions	10
Chapter 4: Demographic surveillance processes	
Chapter 5: Development and Environmental Surveillance using GIS	30
Chapter 6: Data Management	40
Chapter 7: Ethics Considerations	47
Chapter 8: SOMAARTH DDESS Administration	

Purpose of the Operational Manual

This operational manual is intended for use by two audiences:

- those who plan to establish a demographic, development and environmental surveillance site (DDESS), and
- those who have already established a functioning site and are now looking forward to introduce new themes in their surveillance system.

Preface

Following a MOU with the Government of Haryana, SOMAARTH-DDESS Palwal was established by The INCLEN Trust International in 2009 to undertake community-based studies on health and allied disciplines. Adopting a catalytic health system strengthening approach, INCLEN is closely engaged with the community of 51 contiguous villages (SOMAARTH villages) in Palwal, Hathin and Hodal blocks of Palwal District. The DDESS encompasses about 200000 individuals from different religion and diverse socio-economic background. The area is geo-spatially coded and an updated census is maintained for surveillance using latest research and IT tools.

The SOMAARTH Palwal experience is being leveraged to establish two other DDESS: SOMAARTH-Barielly (Uttar Pradesh) and SOMAARTH-Shillong (Meghalaya). This manual will be a useful resource for establishing and facilitating operations at the SOMAARTH sites.

The manual covers following aspects:

- 1. Background for developing a community site.
- 2. Stepwise processes in developing the site
- 3. Detailed operational methodology following the establishment of site
- 4. Benefits and challenges in establishment of the community site

Chapter 1: Introduction

SOMAARTH



"SOMAARTH" is derived from two Sanskrit words: "Som" means highest form of physical, mental, and spiritual health. "Arth" means money, wealth and resources. SOMAARTH ideology, envisions synergy between economic development, social changes and health of the individuals, families and communities.

The SOMAARTH Demographic, Development and Environmental Surveillance Sites (DDESS) have the following aims:

- Establish an integrated surveillance system that is simple, feasible, affordable and sustainable, addressing health in the context of environment.
- In the background of "Whole of Society" transition, determine the interplay between economic environment, agriculture, markets, urbanization, town planning, built environment, education, globalization and socio cultural factors that influences population behavior and its impact on biology, nutrition, physical and psychological health of individuals, families and communities.
- Provide a platform for multi-disciplinary, policy and program relevant studies with multisectoral collaboration the same geographic area.
- Establish an Integrated Management Information Systems (IMIS) to facilitate the collection, analysis, reporting, presentation, and use of data at local, district, regional, and national levels for the overall aim of bringing convergence between development and health.
- Develop pragmatic models of synergistic development with defined health outcomes in collaboration with communities and scaling up to other areas.
- Explore innovative models of Public-Private Partnership for surveillance mechanism and primary health care service delivery.
- Evolve SOMAARTH in to a platform of multi-disciplinary learning and capacity building for next generation of health and non-health academic leadership.

SOMAARTH has the following surveillance components:

Demographic Surveillance: It encompasses periodic documentation of the size, structure and distribution of populations in the area, and spatial and/or temporal changes in them in response to basic demographic events (birth, migrationand death).

Development and built environment surveillance: It covers documentation of spatio- temporal changes in land use (including commercial, industrial, institutional, educational, transportation) and contextual structures.

Environmental Surveillance: It includes surveillance of the indoor and outdoor air quality, ambient metereological data i.e. temperature, humidity and wind direction as well as water and sanitation.

<u>Surveillance of community health</u>: SOMAARTH sites have the objective of strengthening the public health system and related community processes through community engagement, needs assessment, epidemiological data collection and close partnership with the Government health system.

Box 1.1: Geo-demographic details of SOMAARTH DDESS- Palwal

Palwal, Haryana (India): Haryana is a relatively prosperous state in North India surrounding Delhi NCR.. Palwal is a newly created district of Haryana neighboring Faridabad and Mewat districts. The district headquarters, Palwal town, lies 30 km beyond Faridabad on the Faridabad-Agra national highway. When created, the district had four blocks-Palwal, Hodal, Hathin and Hassanpur with 282 villages and area of 1,359 sq.km. Total population of Palwal is 1,042,708 (decadal growth: 25.8%; 554,497 males, 488,211 females). Sex Ratio of Palwal District is now 880 while child sex ratio (0-6) is 874. Density of Palwal District per square km is 767 (2001: 606 per sq.km). Average Literacy rate for Palwal District is 69.32 percent. The districts has primarily an agriculture driven economy but is experiencing rapid industrialization and urbanization.



SOMAARTH DDESS (Latitude: 27°53'59.46"N to 28°7'30.02"N; Longitude:

77°10'2.95"E to 77°22'47.35"E): With an area of 251.7 Km Sq, it encompasses 51 contiguous villages (1,99,702 rural population) from three blocks of Palwal (Hathin, Hodal and Palwal) that are circumscribed by three major roads i.e. – Mathura national highway (NH2) that forms the sites's eastern boundary; Palwal–Hathin state major district road (MDR 135), forming the western boundary; and connecting major district road (MDR132) between Hodal and Nuh forming the southern boundary. This site has a demographic advantage of having a mix of populations from different religions, socioeconomic and overall development status. This site has been selected based on assumptions of rapid economic transition and emergence of business and educational activities that will lead to changes in behavior and health of the individuals and the population at large. A peripheral express highway (Kundli-Manesar-Palwal or KMP expressway or western peripheral expressway) is fast coming up and which traverses through some part of the site. Additionally, special economic zones (SEZs) are declared on either side of the proposed expressway to boost industrial growth and business. Along NH2, one of the busiest roads in the country, several major industries, educational institutions and hospitality businesses have come up over the last three to five years. The real estate has gone up by more than five folds in anticipation of the growth potential of the area.

On the basis of developmental attributes and access to the highway we have divided the surveillance area, broadly into three regions:

- (1) Villages within 4 kms of the proposed KMP Highway (n-15): The eastern and western peripheral expressways (Kundali Manesar Palwal [KMP] and Kundali Ghaziabad Palwal [KGP]) shall traverse the region; and several SEZs with new manufacturing and service industries are likely to be established in the area.
- (2) *Villages on existing NH-2 Delhi-Agra Highway (n-12):* Villages within one km of Agra highway are experiencing change over a considerable period of time. Several land stretches are declared as educational, commercial and industrial lands wherein several institutions and business establishments have already come up, in addition to private technical and management institutions;
- (3) *Villages relatively far from these two highways (n-24):* The south-west part of surveillance site belongs to the previous Mewat region, one of the most backward area in the North India and predominantly populated by Muslims. This area is challenged with low literacy, economic status, health indices, and other development indices.

Chapter 2: Initial Work-up while establishing the DDESS

- Knowing the area:
 - **Preparatory work**: Conducting background study i.e. understanding the area and resources using census of India, National Health Family Survey, District Level Health Survey, Annual Health Survey, administrative documents etc.), identifying potential stakeholder constituencies. An internal communication brief should be written and shared with all the team members in order to have consistency in the content of the dialogue. Based on the secondary data (district statistical handbook, Bhuwan etc) potential periphery demarcation should be done. Budget and timeline should be planned.

• Community engagement

- Community engagement is one of the most important and continuous process. For each of the sub activity tools and techniques need to be prepared beforehand.
- o Area visit
 - Keeping background information of the area, about the organization, one pager on the proposed plan and internal communication brief.
 - Identification and interaction with key stakeholders (local administration i.e. district – block level, Panchayati Raj Institution (PRI) members, other agencies having community presence)
 - Assess political, social, economic and religious environment
 - Transect walk in the villages and informal interaction with the community.
- Interaction with community
 - Meeting with village leaders and community members.
 - Group discussion and one to one semi-structured interactions with the community leaders.
- o Desk work
 - Summarized report about the visit with inputs from all the members of field visit team and getting feedback.
 - Mapping of agencies' presence in the geographical area and based on the inputs gained during the field visit, refinement of the surveillance site periphery.

• Administrative procedure

- Permissions: Seek administrative approvals from competent authority to work in a particular area (Drafting of MoUs, taking permissions etc.)
- Setting up field office:
 - Identification of location: Field office should have good access and a neutral location i.e.
 should not have influence of any political, religious or administrative parties or person. The

location of the office should be such where anybody /or everybody from all the surveillance villages can visit with ease.

- Logistics and infrastructure arrangement: Office setup should match the local area SES attributes.
- Staff recruitment: Preparation of a manpower requisition document highlighting number of staff required, qualification / experience and job responsibilities and timeline to have them on board. Team should be planned as per the project need. The core surveillance team could have following members- Project PI/ site in charge, station manager, research officer census, research officer- GIS, GIS analysts, community coordinator, field workers.
- Development of manuals
 - Preparation/ refinement of operational manual
 - Preparation of training manual (procedural manual)
- Development of data management infrastructure
 - **Design, pilot testing and finalization** of project specific data collection tools
 - **Design, pilot testing and finalization of** electronic software, equipment, forms and processes
 - Understanding the harmonization of surveillance data and project specific data collected from the field.

Chapter 3: DDESS: Key Definitions

SOMARRTH DDESS is a comprehensive surveillance platform as it includes three domains of surveillance i.e. – Demographic, Development and Environmental surveillance. The focus of demographic surveillance is individual and households; development surveillance focuses on the land use and land cover; environmental surveillance focuses on monitoring of physical environment including weather conditions within the surveillance site. SOMAARTH DDESS design, core concepts and definition pertaining these three main focus areas are described in this chapter.

Every demographic surveillance system is required to define the population under surveillance. As most individuals within any population have places of residence and attachments to social groups, the task of defining the population begins with the identification of the land parcel, households (where applicable), and individuals that will be visited and observed and assigning them unique codified identifier numbers that can be tracked and followed-up for identifying temporal changes [Box 3].

	Box 3.1: SOMAARTH Key Definitions
Land Parcel	A piece of land including all physical structures built on it (if any), has at least one owner and is being used as residential/non-residential or lying vacant or has ruins/under construction is called land parcel. A land parcel has clear boundary and has its own shape and size. In rural areas a land parcel may not always have a demarcated compound wall or fencing. Sometimes boundary can also be demarcated by hedges or planting. This piece of land is not available for agricultural use.
	Types of Land Parcel
	 Residential land parcel: A land parcel which is exclusively used for residential purpose. Residential land parcel should have a physical structure or building which has been occupied for residential purposes. Non-residential land parcel: A land parcel which is exclusively used for non-residential purposes. Non-residential land parcel should have at least one structure or building which has been used for non-residential purpose. Example: for shops, cattle sheds, baithak, pump house, clinic, etc. Mixed Land Parcels: Land parcel having both residential and non-residential activities. A mixed land parcel is comprised of minimum one residential and one non-residential unit which may be on same or different floors. Mixed land parcels are very common in rural areas and they should also have at least one building or
	physical structure within its boundary.
	• Vacant land parcel: A vacant plot without any structure built on it. It may or may not have a demarcated boundary in rural areas of India.
Physical	A physical structure (of any shape and size) built within the territory of any land parcel
Structures	which usually has four walls and a roof. There can be one or more physical structures on a single land parcel
	On the basis of construction materials used, physical structures are classified in
	following three categories:
	1. <u>Pukka:</u> A pukka structure is defined according to the composition of their wall and roof. The wall materials should include bricks, cement bricks, concrete, stones

	 and GI/Metals /asbestos sheets. The roof materials should be concrete, machine made tiles, burnt bricks, cement bricks, stones, slate, GI/Metals /asbestos sheets. Kuchcha: A kuchcha structure with walls and /or roof which is predominantly made of materials other than those mentioned in pukka structure. The wall materials should include unburnt bricks, bamboos, mud, grass, reeds, thatch, and plastic, polythene, loosely packed stones. The roof materials should be unburnt bricks, bamboos, grass, reeds, thatch, plastic, polythene. Mixed: When either wall or roof of a structure is made of Kuchcha or Pukka materials as mentioned above.
Unit	On the basis of activities, physical structures are divided into residential unit or non
	residential unit. Physical structure used for residential purpose is known as
	'Residential unit' or household and structure used for non-residential purpose is
Decidential	Known as Non-residential unit.
unit or	kitchen is termed as Household
Household	1 Individuals from a household may or may not be related (neither blood
mousenoid	relations nor conjugal relations).
	2. Number of individuals who belong to a particular household may vary from
	single individual to multiple.
	3. Following exceptional situations will also be considered as households:
	- Related / Unrelated individual(s) living under the same roof but
	getting food from outside.
	I ney may or may not nave a kitchen in the dwelling unit. Household are also known
Type of	1 Normal household: A group of persons (related/unrelated/mix) living together
Household	under same roof of typical land parcel and sharing common kitchen.
	2. <u>Institutional household</u> : A group of individuals (related/unrelated persons)
	who live in an institutional buildings and take meals from a common kitchen
	is called Institutional household, Considerations:
	a. Paying guest accommodation: group of people(mostly unrelated)
	living together under same root and taking food from a common
	category
	**hostels are guest houses are considered under non-residential
	categories.
	Homeless household: Households which don't live under the roof, but on roadside,
	under flyover/pipes/staircases/platform/bus stop are termed as homeless household
	and these are not currently registered under SOMAARTH DDESS.
Non	A physical structure used for non residential activity are called non residential units.
residential	Based on the main activity non residential units are divided into 10 broad categories.
Fligible	Eligibility criteria provided in chapter 3
Individual	Englointy enteria provided in enapter 5.
Note: Housel	hold is a group concept which is accounted on individuals and how they are living.
Household s	taying for at least 6 month in a village under SOMAARTH DDESS will only be

considered eligible for registration.





Giving numbers/Location Identification numbers

Numbering the Sector: The numbering process evolved through sub dividing villages into sectors (chapter 4 secotization process). Sectorization followed natural or manmade linear features (road, canal, drain etc) as a boundary of demarcation. All the sectors were assigned ID with English alphabets i.e. A, B, C, and D.....and so on. However, attempts have to be made to sectorize villages following their basic shape and size i.e. in villages with concentric settlement pattern, numbering of the sectors will be done in clockwise spiral pattern while villages having linear pattern will be sectorized using linear numbering pattern. Sectors were made up of minimum 1 to maximum 600 congruous land parcels.

Numbering the Land parcel: Arabic numeral to be used to number land parcels. For each land parcel unit, the assigned number will be a unique one to that sector and won't be assigned to any other land parcel unit. Any land parcel unit number will be expressed with its 'sector code' prefixed to the number. For example, A003 means 3rd land parcel unit of sector A of a study village. Similarly B034 means 34th land parcel unit of the sector B of a study village.

- 1. Land parcels were given number using left hand rules.
- 2. Land parcel identification number is dependent on ownership, proximity and usage all three criteria were checked while giving land parcel identification number. E.g. If R or NR unit of a person are not congruous then different IDs were allocated.
- It was found that land parcels have multiple units spreading both horizontally and vertically. The subunits were also mapped in GIS maps using hypothetical lines to provide depiction in the 2D map.
- 4. All the structures in the built up area were covered using UIDs even public toilets, chaupal, temples etc were given land parcel identification number.

- 5. Land marks, vacant spaces, canal, wells, roads and other environmental datasets were stored separately in spatial database.
- 6. Shops were given separate UIDs but if the shop is operating in very temporary manner without any specific infrastructure they were given sub unit identification number.
- 7. Destroyed structures, ruins and land parcel under construction were given numbers.

Numbering the split or merger: Splitting and merging of units in land parcel happens during the update round when two or more units (residential or non residential) become one unit (merging) or one unit (residential or non residential) splits into two or more units (splitting). However, the splitting and merging of units happen over time/years and hence will not be captured during baseline survey. These scenarios will occur during update survey in the surveillance site. Split/merger number will be "00" in the baseline survey and will be automatically updated in the database.

Numbering the unit: All the residential or non residential activities taking place within a land parcel will be caprured through the residential and non residential forms. Each unique activity covered is called unit. Units within a land parcel will also be numbered with Arabic numerals but with two digits i.e. 01, 02, 03.....99 (this is simply because, units at maximum, though theoretically have 99 structures). In standalone (not in household enumeration form) unit number will be expressed in "()" that will be prefixed by Land Parcel ID where it resides. For example household No A 003/ 1 means, in sector A, 3rd land parcel and 1st structure (R/NR) in that land parcel.

Other points considered while a Location ID is allotted to the Land parcel:

- a. For each household, the assigned number will be a unique one to that sector and won't be assigned to any other household of that sector of study village.
- b. Structures, that are owned by a household but not situated within the same premise, or separated by road/public places (e.g.- Baithak, temple, another house), must be treated as separate land parcel (s) and must be given separate land parcel ID
- c. Units IDs (prefixed by Land Parcel IDs) will be allotted to sub units.
- d. If a new house is constructed after accomplishment of numbering of all land parcels and in a village, the new construction will be numbered with a new numeral which must be in continuum with the last land parcel of the respective sector, even if it is geographically/physically located far off in that sector and not in sequencing. Suppose last land parcel number of a village is A 045. Now a new construction has come up between land parcel number A015 and land parcel number A016. This shall be numbered as land parcel A046.

e. In case of multistoried structure with multiple R/NR units, different floors shall be considered as different units if they fulfill the criteria to be called as units. In this case noconsideration for their gates will be given to call them as a seprate land parcel. irrespective of their gates. These shall be numbered with a "/" prefixed by land parcel IDs

SOMAARTH surveillance site is experiencing rapid structural and economic development. Thus, for next few years there will be construction of new structures as well as demolition of structures. To introduce dynamicity to the whole numbering process, following protocol has been adopted.

1. The land parcel unit numbering will be done as follows:



Unique Identity of individuals:

Unique identifiers to every individual members residing in SOMAARTH DDESS is established by linking the subjects in a land parcel unit through a hierarchical system of unique numbers. These are built up from a unique number for the parcel unit, followed by serial numbers for each of the households within it (where the notion of households applies) and then for each of the enumerated individuals within each household. The unique ID for land parcels unit is in turn generated from a composite number assigned to Block, village, sector and the land parcel unit number.

In this system, the unique number for each individual in the DDESS is a composite of the numbers for the **block, village, sector, land parcel unit, and household member.** This involves creating complex hierarchies, in which the unique number of the residential unit itself is a composite reflecting allocation to regions, areas, and villages (where they exist). These IDs are unique for individuals. As individual moves from one land parcel unit to another the individual Id gets change. Thus an individual in his lifetime may have several IDs as he moves from one dwelling unit to other but these are only unique to him/her. In the data system multiple IDs in different time period depict the spatial movement of the individual.



Chapter 4: Demographic surveillance processes

Defining SOMAARTH Surveillance System

SOMAARTH surveillance system is defined as a set of field and computing operations to handle the longitudinal follow-up of well-defined entities or primary subjects (Demography: individuals, households, family, development: non residential land use, Environmental: physical environment (covering residential, non residential and vacant land parcels) and weather conditions) and all related demographic and health outcomes within a clearly circumscribed geographic area (51 villages of Palwal district). Unlike a cohort study, a DDSS follows up the entire population of such a geographic area.

An initial mapping (with geo-referencing) and census will define and registers the target population. Regular subsequent rounds of data collection at prescribed intervals will make it possible to register all new individuals, households, and land parcel changes and to update key variables and attributes of existing subjects. This will form the core system for SOMAARTH DDESS.

Thus, core system shall provide monitoring of population dynamics through routine collection and processing of information on births, deaths, marriage and migrations, change in the land ownership, change in the head of the households— the only demographic events leading to any change in the initial size of the resident population. This core system shall be complemented by various other data sets that provide important social and economic correlates of population and health dynamics. In addition to this a routine (in every two years) update of Geo referred attributes for physical environment (including built environment) will be done under this core system through wave based data collection activity.

At SOMAARTH, direct method (first-hand enumeration, verification and measurement; survey/ census approach) for data collection is observed. It is divided into two broad categories:

- 1. Core Baseline Round
- 2. Update Rounds

This chapter discusses the standard practices followed during data collection.

Baseline round

Purpose: Listing and mapping activity; the initial (baseline) census enumeration.

The initial census focuses on obtaining information across three broad domains in align with the SOMAARTH philosophy:

- Demographic surveillance
- Developmental surveillance
- Environmental surveillance

The process of data collection is described in the next chapter. The information collected under these broad domains will then be used to derive various indicators for mapping transition at all levels. The core variables which are to be documented at the land parcel (LP) and unit level in the surveillance area include the following:

Demographic surveillance	Development surveillance	Environmental surveillance
-Use of unit-Residential/	-Ownership of land	-Land parcel and unit construction
non	-Family income – contribution by all	characteristics -Source of light and
residential/vacant/ruined/use	family members	ventilation -Cooking practices-Water
not clear	-Socio economic status	availability & uses
-Details of land parcel owner	-Domestic animals and Assets owned	-Toilet facilities
-Details of respondent	by the household	-Sanitation practices
-Details of unit head	-Economic environment at community	-Disposable waste management
-Demographic details of the family	level (Non residential)	
-Education and occupation	-Mental/ Physical Disability	
-Population pyramid-Family structure	-Consumption patterns-Tobacco,	
	alcohol, substance use	
	-Chronic disease status	
	-Use of health facilities al)	

Table 4.1: List of Core variables documented at the land parcel and unit level

Another set of information to be obtained at the time of baseline census is the mapping of temporary/migratory population. This population is most vulnerable and is left out from all the interventions taking place in that area due to their mobile nature and temporariness. The area where these populations are residing has to be marked on GIS map and has to be assigned a temporary ID.

The other set of information like lifestyle behavior, immunization schedule etc. shall get captured in the branched out research projects of the surveillance sites. The broad steps leading to the enumeration of the cohort characteristics at baseline are as follows:

Membership of the SOMAARTH DDESS and Residency status

Eligibility as SOMAARTH DDESS member (This primarily deals with whom we should include as part of surveillance)

- As land parcels have fixed geographical positions, there are consistent and simple rules for their inclusion: they are included if they are situated within the delimited boundary of the SOMAARTH DDESS.
- At baseline all households are eligible if (and while) they are located in the villages under SOMAARTH DDESS.

- At baseline round all individuals residing in SOMAARTH surveillance area will be listed as members.
- During update rounds in-migrating individuals (from outside DESS area) are only eligible once they complete a period of **six month stay** in that household.
- During updates all birth and marriages must be listed as members (thus threshold level of duration of stay is not applicable to them)
- During updates if a DDESS member is found to be dead, his/her membership shall be ended.

If an individual is not member of that household for last 6 months he/she will not be considered as eligible for that residence. Thus his/her residency status doesn't change and so the unique identity. It is not true for the birth or for marriage at that household. Conversely, the non-continuous presence of an individual, with short periods of absence, may be considered continuous residency if he or she meets a threshold for inclusion *i.e.* six months of period.

Thereafter, new membership may commence in-migrations exceeding a prescribed threshold of duration, and current membership may end because of deaths or out-migrations (again exceeding a prescribed threshold of duration). New membership episodes may commence as a result of events that initiate a social relationship with a household, such as birth, marriage, adoption, or household formation, and may be terminated by events that end such a relationship, such as death, divorce, or household dissolution.

Deliverables:

- 1. Sector-wise digitized Map of the all villages with all land parcel with their respective IDs
- 2. List of the all units and their IDs
- 3. Census enumeration data of all land parcel and their respective units

Box 4.1:Experience from SOMAARTH, Palwal:

- The baseline census enumeration at SOMAARTH, Palwal was done from Jun 2012 to May 2014.
- *Team characteristics:* The team involved in the activity included SOMAARTH Chief, Station Manager, Senior Program Officer, Technical Manager, Program Officer (Assistant Research Officer), Field Team Leader, Field Supervisor, Field Worker). The technical capacities of the team ranged from Public health, Social sciences, GIS, IT, Anthropology, Medicine.
- Activity quantum:

The data collection was done on paper. It took 20-35 minutes to fill Residential form, 10-15 minutes for Non-Residential form Prior verbal consent was taken from respective LP/unit. On an average 2 to 5 minutes were devoted to get consent.



Figure 4.1: Android data collection process

Longitudinality of SOMAARTH surveillance system:

Longitudinal measurement of demographic, development and environmental variables is one of the key agenda of a SOMAARTH DDESS. This will be achieved through repeated visits at more or less regular intervals to all houses to collect a prescribed set of attribute data on registered subjects, who are consistently and uniquely identified. A regular update of all land parcel attributes (Built environment) and other physical environmental features along with updating of meteorological data shall also provide dynamic follow up of environmental conditions. Repeated recording of events affecting study subjects during the interval between visits allow one to construct their history and differentiate DDESS data from data collected in multi-round surveys and other prospective studies that allow comparison over time only on an aggregated level.

For SOMAARTH core surveillance system, after baseline round, following categories of updation is planned.

- Real time Updation: As soon as a village completes its baseline round, it enters into updation round. The events that require real-time updation are: Death, Birth, Migration and Pregnancy. This will be documented as and when it happens in SOMAARTH and shall be updated with core database.
- 2. Yearly Updation: SOMAARTH plans for yearly updation of, change in land parcel status, change in head of household, in-migartion (household or family), out migration (household or family), marriage. This shall be updated annually. During this update, updation of land use changes shall also be done. Meteorological data shall be updated from both available primary and secondary data sources annually.
- 3. SOMAARTH data collection wave: During the wave data collection phase full census of all the households and non-residential units along with the updates of environmental data will be conducted. All changes in built environment and physical environment will be documented in form of geospatial database.

Update round(s)

Surveillance Life Cycle:

Each census Round will be followed by series of update rounds in order to routinely update the existing surveillance data. Only the major events will be updated during the update rounds (*refer point 3 & 4*). In order to update the complete surveillance data, i.e. the census information, new census round will be launched followed by its corresponding update rounds as represented in the below diagram.

Description of Events:

Events are described as the episodes which have occurred in the Land Parcel / Unit since the last visit (census round). Such events are broadly classified as follows:

Figure 4.2: Surveillance life cycle



 Table 4.2: List of events to be documented in Update rounds

Land parcel Events	Unit Events	Demograph ic Events	Developme nt Events	Economic Event
-	-Unit use	-In-	-Increase or	-Change in
Ownershi	change	migration	decrease in	primary
p change	-Ownership	-Change of	total number	activity of
-Use	change	HOH	of primary	NR unit
change	-Unit head	-Birth	and support	
	change	registration	structures	
		-Marital		
		status		
		change		
		(Marriage/		
		Separation)		
		-Death		
		-Out-		
		migration		
		-Pregnancy		
		registration		

A. Land Parcel Events: Any change in the land parcel characteristics is registered through "Land Parcel Update Form". Following are the characteristics which will be verified for change in each Land Parcel during update round:

- i. Change in ownership
- ii. Change in use type
- iii. Increase or decrease in number of primary structures/support structures

		.5. Details of the failu parter e	venus
S.No	Scenarios	Change in Existing Land parcel	New registration*
Ownership			
1	LP has been sold/purchase d	Update new owner details	Change ownership
2	LP-Property transfer	Update new owner details	Change ownership
3	LP-Death of owner	Update new owner details	Change ownership
4	LP-Splitting	Specify the reason of splitting-sell/purchase; Family division; Death of owner	Register new land parcel and update old land parcel
5	LP-Merging	Specify the reason of splitting-sell/purchase; Family reunion; Death of owner	Merge two or more land parcel /Outmigrate the members of old land parcel in the merged LP/Close the LP which has got merged in the new LP
Use type ch	ange		
1	Residential to vacant/ruined/ Incomplete/ Under construction	Out migration of members	Out-migrate all the members residing in LP and close R unit in the LP
2	Non Residential to vacant/ruined/ incomplete/ under construction		Close the NR unit in the LP
3	Residential to Non Residential	Out-migrate all the members residing in LP	Close the unit and register new NR unit in the LP
4	Non Residential to Residential		Close NR unit in the LP, register R unit in the LP, in- migrate all members in the R unit of LP

Table 4.3:	Details	of	the	land	parcel	events
1 abic 4.5.	Detans	UI	un	Iunu	parcer	C V CHILD

B. Unit level events: Any change in the unit will be done through "Unit Update form". The

following events will be captured at the unit level:

- i. Unit use change event
- ii. Ownership change event
- iii. Unit head change event

S.No.	Scenarios	Change in Existing	New registration*
		Unit	
Ownership			
1	Change in ownership		Change the owner details
	details		of the unit
Unit head ch	ange		
1	Change in unit head		Change the unit head details (in case the tenant is staying in the unit then in that case unit head will be the tenant and unit owner will be different)
Unit use cha	nge		

Table 4.4: Details of Unit level events

1	Residential to Non residential	Out migrate the members from the unit, close the current R unit	Generate new unit id for NR unit and update the activity type
2	Non residential to Residential	Close the current NR unit, generate new unit id for R unit, in migrate the members in new unit	Update the events for the residential unit

- C. **Demographic events:** These events are termed as *"vital events"* and will be registered through pre-defined forms (attached as annexure), which will be filled using Android based mobile application in Tablets. All the forms are interlinked to each other so Field Worker needs to complete all the forms for any unit in order to further process the forms for saving and synchronization. These events include the following:
 - In-migration: It refers to the migration of registered member(s) of the SOMAARTH from his/her current unit to join another unit or form a new unit within SOMAARTH DDESS area. The definition for the residency period is "if a family/its members are living for ≥6 months then only it will be registered in the surveillance area and will be assigned a location and hhid. The in-migration could happen due to the following reasons: work, education, marriage, splitting of unit, political, epidemic, family division. In addition, the in-migration process could be from within the SOMAARTH DDESS area or outside the SOMAARTH DDESS area.
 - *ii. Change of HOH:* It refers to the document the process wherein, the responsibilities of the residential unit are given to some other family member. This event embarks the change in decision making capacity of the unit which has further implication on the health, wealth and education of family members. This could happen due to family division, migration, death, and family dynamics. There is no time cut off for the event. This event will also result in the change of the relationship with the head of household. The android system is capacitated with the feature which automates the process of change in genealogy for all family members.
 - iii. Birth registration: A birth registration form is used to register a pregnancy outcome (live birth/still birth/aborted) for women who are members of SOMAARTH. Live birth is defined as a pregnancy outcome that showed signs of life, such as movement or crying, after being delivered. Still birth is defined as the death or loss of a baby before or during delivery. Both miscarriage and stillbirth describe pregnancy loss, but they differ according to when the loss occurs. Abortion is defined as ending of pregnancy by removing an embryo or fetus before 28 weeks with the consent of the mother.

iv. Marital status change (Marriage/Separation): Marriage is a critical event for SOMAARTH DDESS because it is associated with various other vital events such as changes in demography of household as well as village, migration, and marital status. It also results in other events like pregnancy and birth. Marital status shall be measured in two ways,

a. As status of an individual at the time of first time registration

- b. As an event after initial registration.
- c. During update round, all new marriages are documented through *Marital Status Change Form* and new member shall be included/ excluded through migration processes.
- d. Once individuals' marital status has been registered during baseline or in subsequent rounds, any change should be registered as an event. The event could be classified as: married, living together/cohabiting, divorced, separated, reconciliation, or widowed.
- e. If after divorce the couple is staying together, then it will be captured in WAVE.
- f. Definition of separation: If the couple is staying apart from each other for more than 6 months.
- v. *Death:* The Death Registration Form will be used to register all deaths (except still births) that occur under SOMAARTH DDESS. Effort must be made to obtain as accurate as possible the date of death. For infants, accurate day date information must be obtained. Deaths to non-DSS members such as visitors shall not be recorded. A verbal autopsy will also be done by the research professionals on suitable time after death. An effort will also be made to upload the death certificate of deceased individual. The cause of death should be categorized according to the causes defined by ICD10. The codes for various reasons should be adopted from the ICD 10 to maintain synonymity in the data.
- vi. *Out-migration:* It refers to the migration of registered member(s) of the SOMAARTH from his/her current unit to join another unit or form a new unit within SOMAARTH DDESS area or outside. The definition for the residency period is "if a family/its members have moved out of the unit for ≥6 months then only it will be registered in the surveillance area and will be assigned a location and hhid. The outmigration could happen due to the following reasons: work, education, marriage, splitting of unit, political, epidemic, family division. Note: During pregnancy some women stay at their father's place for 3-5 months or more. This should not be considered as migration.

vii. **Pregnancy registration:** Pregnancy registration is important for improving estimates of maternal and perinatal mortality since outcomes can be deliberately followed. Monitoring pregnancy helps in the early registration of births but most importantly in capturing stillbirths, abortions and neonatal deaths. Pregnancy monitoring starts with the registration of the pregnancy. This will be done during the weekly visits of the field workers, where they will ask their respondents if any woman has been pregnant in the household since the last visit. If there is a pregnant woman in the household, she is interviewed and information about the pregnancy is collected through a *Pregnancy* Registration Form. When these pregnancies are registered, data entry takes place and monthly lists of all pregnant women by clusters (The villages under each fieldworker) will be produced. FWs shall move with these lists during their normal rounds and for each household where there is a pregnant woman on the list, the FWs ask to know the state or outcome of the pregnancy. Depending on the outcome of the pregnancy as seen at an update round, a pregnancy outcome form will be filled. Where the outcome is a live birth, it is registered and also the birth registration form is filled.





- D. **Development event:** This event will capture the change in number of primary or support structures in the land parcel.
- E. **Economic event:** This event will capture the change in the primary activity of the non residential unit.

Data Collection in Update Round:

The data for the update round will be collected using paperless data collection system (Android Operating system software name: "SOMAARTH 3"). The major objective of the update round is to document the major events which have taken place at the level of land parcel and unit. This would require a network creation and community mobilization by the field team. This network would consist of key informants including – ASHA Workers, Aanganwadi Workers, Helpers, ANMs, religious leaders and panchayat members in order to track the new and missed events of his/her assigned surveillance area.

Data Collection for the update round will be done under following pre-defined steps in the mobile based software by the field team:

- Step 1: Update Land Parcel or register new land parcel(refer point 4A)
- Step 2: Update units in a land parcel or register new units (refer point 4B)
- Step 3: Update events at the unit level (refer point 4C)

**the android device of the field worker will have a listing of land parcels according to the assigned sector in the assigned village. As a first step, the field worker will be syncing his/her android device with the server which will update the line list of land parcel and their respective units in his/her device database.

Data Synchronization

Data will be filled offline in the field which will be synchronized with the online server once the worker comes under internet connectivity zone. SOMAARTH Mobile application will have synchronization button on home screen of each user which s/he has to press to synchronize the data between tablet and the server. The main objectives of data synchronization will be to update the list of tasks required by field worker/team leader to further progress with their work.

- o List of sector assigned to the field worker
- List of all registered land parcel in the database within that sector of a particular village
- List of all outstanding cases including out migration, refusal, locked LP/unit, respondent not available, use not clear type of LP

Quality Control

Batch processing:

The quality control (QC) protocol is executed on a randomly selected sub-sample of forms filled in the by Field Workers (FW). These forms are selected after creating batches of forms. The steps followed are as below:

- 1. Create batch of 20 land parcels and its associated forms via automated process. The system will stop synchronization of data between some system defined time intervals in the night and will create batches of Land Parcel forms. The system will create batches of 20 Land Parcels forms (configurable in configuration file) filled by one Field Worker. Batches will be created based on Land Parcels and FW-wise. One batch will have all the forms filled by the same Field Worker. If forms are not in the multiple of 20, Data Manager can create batches manually for those remaining forms. For example, if there are 48 forms filled by FW1 and 77 forms filled by FW2 at the time when system stops synchronization. Then, the system will take 40 forms filled by FW1 (create 2 batch) and 60 forms filled by FW2 (create 3 batch) for automatic batch creation.
- 2. From this the 20% record of a batch i.e. 4 LP forms will be picked for QC process. Among them 2 records with changes and 2 records without changes will be selected or any four records will be selected if the combination is not possible.
- 3. Assign 50% of forms (2 forms) of a batch, which have 1 not changed land parcel and with 1 changed land parcel to web QA1 for verifying through listening audio files. Web QA1 will do the voice verification for the records and based on that QA1 will accept or reject the record.
- 4. Assign 50% of forms (2 forms) of a batch, which have 1 not changed land parcel and 1 changed Land parcel to Android QA2 for re-collection of data. Android QA2 will collect and fill the data again for the field.
- 5. Data Manager views the status of pass/fail done by Android QA1.
- 6. Data Manager compares the forms and data filled by Android QA2 with the respective land parcel filled by Field Worker.
- 7. Data Manager accepts or rejects the batch. DM can accept/reject a batch only if there is no pending form in the batch. If there is some form, where an action is pending, then DM cannot accept/reject the batch.



Figure 4.5: Details of batch processing mechanism

Verify Data with Attached Voice File (QA1):

- QA1 user will validate the details with the help of audio/voice files attached. However, QA1 user can verify the forms of only those Field Workers, which are mapped with him/her.
- If QA1 user finds the form data accurate, he will approved the form. The approved form will be processed accordingly and will be displayed with the Approved status on the screen where DM views the respective batch.
- If QA1 user does not find the form data accurate, he will disapprove the form. The disapproved form will be processed accordingly and will be displayed with the Disapproved status on the screen where DM views the respective batch.

Recollect Data (Android QA2):

Android QA2 will visit the respective site again and collect data for assigned land parcel. However, QA2 user can collect the data of only those Field Workers, which are mapped with him/her.

Accept/Reject Batch (DM):

After analyzing the forms of the batch, Data Manager will either accept or reject the batch. However, Data Manager can perform an action on a batch only if all the forms of that batch have been selected for QC process. Means, the application will not allow Data Manager to accept/reject a batch if there is/are any form(s) in that batch Pending for taking an action by QA1 or by QA2.

Validate Locked and Refusal cases:

Field Team Leader (FTL) visits the locked or refusal cases, which have marked Locked or Refused in the system by the Field Worker. If Field Team Lead is also unable to get the details, then he/she will approve this case as Refused or Locked with comments. Site Admin and Data Manager can view the Locked or Refused cases in Web view. If Field Team Lead receives the details, then he/she will disapprove and Field worker will get notification to collect the data of same cases.

Chapter 5: Development and Environmental Surveillance using GIS

SOMAARTH Geographical Information system (GIS) is one of the main components of Demographic Development, Environmental Surveillance Site (DDESS). SOMAARTH DDESS GIS is innovative in the development of a comprehensive geospatial data through synthesising contextual and compositional elements of the environment and integrating them into the demographic, development and environmental surveillance system.

Aim

- 1. To construct high resolution geospatial data of rural built environment.
- 2. To identify and monitor environmental risk conditions and contextual pathways associated with various communicable and non-communicable diseases.
- 3. To innovate evidence based environmental and behavioural interventions for prevention and control of health problems.

SOMAARTH GIS data variability

- **Built environment:** Built environment or built up area in villages are spatial sprawl where village habitants reside and engage in their livelihoods. Eventhough it coveres the vacant land or plots upcoming for future socio economic activities, it do not covers the agricultural farm land of the villages. SOMAARTH GIS captures and characterizes the developmental (roads, railway, buildings, land use) and environmental components (land cover, waterbodies, waste land) of the rural built environment.
- Land use: Each and every land parcel within the built up area are assigned land use classes. Anderson's international land use classification system is adapted keeping in view the local contextual features and land parcel and adopted at SOMAARTH GIS. The resulting land use classification system include three levels. Level I representing 'Built-Up Land', 'Agricultural Land', 'Water Bodies', 'Waste Land' and 'Forest Land'. For example, Built-Up (Level I) is further refined to Level II, to include the classifications of 'Residential', 'Commercial', 'Industrial', 'Institutional', 'Utilities', and 'Services', 'Transportation', 'Agricultural and Others'. Subsequently, Level II categories are further refined into a Level III classification.

- Weather and air quality: GIS integrated weather and air quality monitoring system is establishment, covering point based recording of the real time ambient air quality (PM2.5) and other meteorological attributes i.e. temperature, humidity and wind direction at one or more locations within the surveillance site. GIS maps are very helpful in slecting the strategic location for establishment of small weather station within the surveillance site. Latitude and longitude information of weather data are used for integrating weather and air quality data within the geospatial datasets. At SOMAARTH DDESS Palwal, the system for PM^{2.5} air quality monitoring at the ambient level is established and upgraded to drone based observations for monitoring the dispersion of particles (PM2.5) at different altitudes and measurement of temperature inversions. The personal exposure monitoring was also carried out in selected female subjects (primary cook) from the site villages.
- Water and sanitation: Water and sanitation are two critical components of the village environment. At SOMAARTH water and sanitation assessments are carried out using high resolution GIS maps depicting the drinking water pipe lines, drainage system (drainage channel and their quality), liquid and solid (litter) waste spots i.e. open litter of the large size (covering more than 1 meter diameter) as well as location of water stagnation and water spilling areas.

Data Construction

In the absence of geospatial administrative datasets from secondary sources, baseline GIS mapping of built environment has to be carried out for establishing a GIS integrated comprehensive surveillance platform. The steps of data construction below are given below (see figure 3).

Step 1 Preparatory activity

Developing the surveillance platform is a long term commitment and require continuous support from the local stakeholders including community members. Stakeholders engagement establish the networks required to later undertake participatory mapping and census processes within the villages.

Three field teams are to be constituted: Census, GIS and Environment teams. Teams are comprised of lead personnel with public health, geography, and environmental science backgrounds; for field staff, local residents with graduate and undergraduate qualifications should be hired. Project personnel are trained through three separate structured 2-week training programs which included classroom sessions (20% time) along with hands on fieldwork (80% training time). A village mapping listing manual, census enumeration guide, and GIS mapping guidelines are prepared to ensure the consistency in data collection processes across the site. Separate micro-plans for collecting data sets pertaining to geospatial, demographic and environmental domains should be prepared and instruments.

Box 5.1: Tools

- Bhuvan TM/ Google EarthTM open source imagery
- Survey of India (SOI) district map at 1:50,000 or 1:25000 scale
- High resolution (<1 m) satellite data
- GIS software such as ESRI Arc Map and Qunatum GIS software
- Air quality monitoring instruments MetOne E-sampler 9800 for ambient PM_{2.5} and meteorological data (wind speed and direction, temperature, relative humidity), UCB-PATS+ for household PM_{2.5}, MAXIM i-buttons for stove usage monitoring and DJI Phantom-1 for recording the particle dispersion and temperature inversions

Box 5.2 : Acquisition of high resolution satellite data

In India National Remote Sensing Centre, Hyderabad, is the nodal agency for selling national or international satellite datasets. Cartosat 2^{TM} (Indian Remote Sensing Satellite), Quickbird TM, World view TM (Digital Globe) Plaides TM (Air Bus) are some of the satellite products suitable for conducting mapping exercise for surveillance purposes. Four- five band bundle (4 multispectral and 1 panchromatic bands) satellite data are good for land parcel level digitization purposes. The acquired images should have very less radiometric noises and almost zero cloud coverage. Pansharpening helps in enhancing the sharpness of the images which supports data digitization process.Due to the security reasons high resolution satellite (HRS) images (<1 m resolution) are not directly sold to private agencies. However private agencies can purchase satellite data via any government agency's endorsement.

Step 2 Participatory mapping and line listing activity

Participatory mapping and line listing activity are the key processes which helps in registration of surveillance area as well as individuals living within that. SOMAARTH DDESS captures individuals as nested in their local geographies. Before starting the field work, team of Research Associates – GIS; prepare the base maps utilizing the survey of India toposheet (scale 1:50, 000). BhuvanTM/ Google EarthTM imagery depicting locations of major roads, water bodies and land parcels. Villages are subdivided into sectors taking roads as a boundary of demarcation. These base maps are supplied to the field team comprising of two field workers (mapper and lister) for each village for further attributation. At first base maps are verified taking community inputs, then field workers mapping listing processes. Field workers identify main entry point of the village on the map, orient themselves as per the directions provided on hard copy of the base map and following left hand rule systematically maps or edit roads, lanes, water bodies, land parcels and landmarks drawn on the maps. These hand drawn maps are called village field drawing. This process of taking community inputs in mapping activity is called participatory mapping. Sector demarcated on the maps are readjusted making each sector comprising of around 50-400 (population of approximately 100-2000 persons) contiguous land parcels in the core

habitation area and 0-50 land parcels in sparsely built outer village. Sectors are given unique alphabetic identification codes in a systematic clock- wise order. Within each sector using the left hand rule all the residential, non-residential, vacant, ruined and permanently locked land parcels are mapped in the form of polygons of relative sizes and shapes of area as informed by the property owner/respondent. Each land parcel charecterised on the map is given a unique location identification number (UID) by prefixing the sector identity and unique numbers in sequential manner following the left hand rule. This systematic approach later helped in developing the location based addresses for each household and non-residential features of the study villages. Line list prepared for each land parcel consists of the following details of structure type as per construction (mud, cement, brick) and usage (residential/ non-residential), ownership, head of household, religion, caste, gender and age composition of household members. The Field Supervisor conductes a 10% random field based checks stratified according to the task accomplished by primary field workers and lot quality assurance approach is adopted for accepting or rejecting the lot.

Box	x 5.3: Development of high resolution administrative data
Village built up area	Boundary demarcation of village built up area becomes a challenging task in the expanding rural areas. As a matter of practice government takes the thoroughfare around the village as the boundary called <i>'phirni'</i> . Phirni are village's built up area boundary demarcation in government land records after the land consolidation. But due to rapid growth of rural areas, there is a mismatch in the boundary existing in government records with reality. In the absence of any updated government records on the village boundaries SOMAARTH mapping relies on the community inputs for the delineation of the village boundary. Mapping focuses within "phirni" of the village where 80% of the residential and other allied land parcels are located. However residential and non-residential parcels located beyond phirni of the village are also captured in mapping exercise in consultation with the community.
Sector	 SOMAARTH coveres the residential, non-residential, ruined, permanently locked and vacant features of built environment. Participatory mapping assists in the subdivision of villages into sectors. Sector coveres around 50-500 (population of approximately 100-2000 persons) contiguous land parcels in the core habitation area. However, outer village sectors are sparsely built and may not have any land parcel. Sectors are given unique alphabetic identification codes in a systematic clock- wise order. Key properties of sectors- Sectors can have upto land parcel up to 400 land parcels. Sector demarcation is done from the main entry of the village from NORTH Direction. Hardcopy of base map with north orientation are provided to fieldworkers. Villages were divided into sectors in order to avoid any disturbances in numbering. The comparatively larger sectors divided by the fieldworkers are reduced in size to an optimum level later by the GIS analysts. The nomenclature of the sectors will be alphabetical e.g., A, B, C, etc.

Land parcel	Within each sector using the left hand rule all the residential, non-
	residential and vacant land parcels were mapped in the form of polygons of
	relative sizes and shapes of area as informed by the property
	owner/respondent. Each land parcel occupied in the residential, non-
	residential or mixed activities was given a unique location identification
	number (UID) by prefixing the sector identity and unique numbers in
	sequential manner following the left hand rule e.g. A001, A002.

Step 3 Geospatial data construction

The analogue participatory maps having contextual details of the study area and line listing assist GIS associates in satellite based digitization processes for constructing digital, georeferenced spatial datasets. Due to the lack of property delineation and informal settlements automated digitization are not possible for rural areas presently. SOMAARTH adopts manual digitization through combining visual interpretation of satellite imagery and participatory maps. Different features are stored as separate feature class i.e. Sector (polygon), roads (polygons, line), water bodies (polygons), land parcels (polygon), well (point), canal and drains (polygon), railway line (line), burial places and landmarks (point). Land parcel features with their location UID and composition data collected during line listing are joined with the GIS layers. Digitization for all the 51 villages is usually done by a team of three to four GIS associates. Program Officer (GIS) conducts random cross checking of 10 % land parcels stratified for every sector in the village for taking corrective steps. Digitised data are projected in Universal Transverse Mercator (UTM) coordinate system.

Box 5.4: Digitization procedure

Digitization rules should be prepared and explained to the GIS associates working on data digitization in order to have uniformity in data extraction. At first GIS associate locates the spatial features (built environment features) made on the field drawings on a satellite image (geo referenced VHRS). After locating the features on the satellite image on screening digitization of roads from satellite image should be done. After complication of road digitaization, sector layer should be created referring the field drawing and sector attributes should be generated. GIS analysts with the advantage of having aerial view resectoris villages using clock-wise concentric ring patterns. Land parcel extraction should be conducted using sector wise systematic extraction approach referring field drawings. Village layers comprise of village sectors, roads, land parcels, water bodies, landmarks and railway line. SOMAARTH village maps include details on features at their finest spatial unit; landmarks were represented as a point location (e.g., temple), residential or non-residential features as a polygon and roads canals as line feature. The polygon representing land parcels were assigned a unique identification (ID) number at the time of the participatory mapping exercise (discussed above). Point locations representing landmarks are updated with names and roads were labelled with their types, quality and surface details. GIS associate should take the ground truthing of data. The volume of spatial coverage and frequency of these ground thruthing field visits varies according to the size of village and difficulty of digitizing satellite imagery.

Step 4 Census data colletion

The hard copies of high-resolution GIS maps and line listing attributes are supplied to census teams to operationalize census of land parcels and related entities. The sector wise high resolution GIS maps (<1:200 scale) facilitate in work allocation and monitoring of census operations. Core variables

collected for the residential structures are: basic land parcel information, demographic details of the inhabitants, household structure, details of construction materials, socioeconomic status, domestic animals and other assets owned by the household, water availability and usage, toilet facilities, sanitation and waste management practices. The self-reported health parameters covered are: details of mental and physical disability, behavioral issues, substance abuse (smoking, alcohol, other substances), health seeking patterns and individuals with chronic disease (illness lasting for more than the last 6 months) in the household. Core variables for non-residential land parcels are: land use typology and waste management besides the structural features and ownership. Regular structured coordination-cumtrouble shooting interaction should occur between GIS and Census teams every week to detect temporal changes and other feedbacks on the maps in real time manner; these meetings help in the regular rectification of both census and GIS data.

GIS data quality assurance

Reliable database are essential for demographic, development and environmental health surveillance system. The geospatial database integrated with demography and environmental data is complex in nature. SOMAARTH uses nested design so that the constructed geospatial database can provide insights about the relationship between place (space) and demography; health. SOMAARTH GIS quality assurance (QA) and Quality check mechanism (Fig 1) ensured good quality of geospatial database. Process standardization and quality assessment were two important pillars of the quality assurance in GIS database. SOMAARTH GIS database design focused more on firmly establishing and following quality assurance mechanism.

Data accuracy elements

- Positional Accuracy: Position accuracy of attributes within a geospatial village data layer, including location of the land parcel, shape and size.
- Attribute Accuracy: Attribute accuracy of land use classifications and road surface types assigned to the geospatial village data layer.
- Temporal Changes: Temporal consistency of data through observing the land parcel changes over a period of time.



В	ox 5.5: GIS data Quality assurance procedures
Process standardization	Database creation across the surveillance site must be carried out following uniform standardized procedures. Task specific procedural manuals are prepared to clearly stating guidelines, duties responsibilities of the different functional team members/ operators and quality objectives of the task. 1. Field worker's manual 2. Census procedural manual 3. GIS procedural manual 3. GIS procedural manual officer. The process adopted for checking and maintenance of data are kept simple and well defined so that they can be understood by all those who involved in the database maintenance. Quality control activity at the specific interval in every phase of the data creation is to be kept. The random errors identified during the quality checking were analyzed so as to identify the roots of the error
Quality Check	 quality checking were analyzed so as to identify the roots of the error and error rectification pathways were drawn accordingly. All the errors and their rectification plan were documented in the manual. SOMAARTH process manuals and repeated quality checking helped in maintaining the integrity and accuracy of data. SOMAARTH DDESS GIS QC plan focuses on mainly 5 components i.e. positional accuracy, completeness, validity, logical consistency and physical consistency. All of these are described below. Positional accuracy: This is checked through site wide ground truthing of map elements. Positional accuracy measured how well each spatial object's position in the database matches with the reality.
	 Completeness: Verification of completeness ensures that entire SOMAARTH DDESS GIS data should be complete in terms of key data elements and procedures i.e. metadata, standard data topology, table structure, precision, projection. Validity: Data validity ensures that each attribute covered in SOMAARTH GIS must comply with defined domain and range. External and internal validity rounds measure the location, attribute and temporal accuracy standards of the database.
	 Logical consistency: Logical consistency is maintained through introducing concept of land parcels. The concept of land parcels synthesizes structures and related attributes in a consistent way. Data verification covered checking of logical consistency in the database. Physical consistency: Physical consistency was checked
	through the visual inspection of the data. SOMAARTH GIS maintained physical consistency through maintaining topological correctness and geographic extent of the database.

Positional quality checking:	It was arduously maintained that all the maps must qualify positional accuracies so that map can represent rural environment as closely as possible. Positional error in geospatial data can be introduced in many ways. Incorrect cartographic interpretation, land parcel extraction, topological errors in point, line and polygons and precision inadequacies are just a few sources of positional inaccuracies. The data production timeline, budget, skills and expertise of personnel involved and budget are some of the other aspects which affect data quality. These errors can be random, systematic, and/or cumulative in nature. Rigorous quality check mechanism helped in maintaining good quality of data. In total three rounds of data verification using the hard-copy plotting of data was used for checking for missing features, misplaced features and registration errors during data construction.
Data acceptance criteria	Certain errors weighted differently than others therefore all the errors were broadly categorized and sources of error were reviewed. All data acceptance criteria were clearly stated in the process manual. Data acceptance and rejection criteria were prepared in advance however initial round of verification had given some more insights about the error typologies in the geospatial database.
On-screen Quality Checking	In addition to on-spot QC at various stages in the data construction the on screen quality check was also adopted. This helped in detecting systematic errors such as an overall shift in the data caused by an unusually high RMS value as well as random errors such as misspelled text. This also helped in maintaining physical accuracy of geospatial database.

Challenges encountered during land parcel extraction and opportunities for data triangulation

i. Issues pertaining to satellite based digitization

In practice the extraction could be constrained at few places due to the spatial heterogeneity and complexities of structures in rural areas. At few places even the high resolution satellite data may not be of much use due to poor spatial resolution. This can be quantified using subjective assessment technique by trained geospatial analysts. Some of the challenges to visual interpretation of satellite data are:

- 1. Multi storey buildings may not be distinguishable.
- 2. Closely packed buildings, connected building and shelters are difficult to distinguish.
- 3. Complex pattern of roof and different building material.
- 4. Varied size, shape and structure of the buildings.
- 5. Poor satellite spectral resolution due to the topology
- 6. Structures partly or fully obscured by vegetations.
- 7. Temporary structures, bamboo and thatched houses.
- 8. Cloud coverage and poor reflectance.

ii. Issues pertaining to field drawing based digitization

Field drawings are more or less qualitative data as at some places these are based on the field workers knowledge, opinion and perception e.g. breadth of roads, landmark location, structure numbering may be marked with varying accuracies. All the inconsistancies between satellite data and field drawings are to be removed through rounds of field verification and village level consultations.

iii. Triangulation

Field drawing contains detailed contextual information but often lacks in locational accuracy. Satellite data helps to get locational accuracy and precision. Identification attributes collected during quantitative survey provides details on ownership and other characteristics. Hence digitization of land parcels should not be only based only on remote analysis and manual digitization of imagery should be done taking constant inputs from field drawing and line listing data. This triangulation approach provides good degree of accuracy in the database.





Chapter 6: Data Management

Comprehensive longitudinal datasets pertaining to demographies, development and environmental components require robust data architecture and advanced information system. This chapter aims at describing three components of data management at SOMAARTH DDESS.

- 1. SOMAARTH data architecture
- 2. SOMAARTH data storage software system
- 3. Core data management processes

SOMAARTH data architecture: A three tier surveillance architecture is established to enable incorporation of demographic, development and environmental domains having capabities for data collection (input layer), data management (application layer) and data harmonization (database layer) [See figure 6.1]. Datasets are prepared to permit relational documentation across each layers and to dynamically integrate additional information from research projects, health facilities and institutional records in a timely manner as datasets are made available to the research team.





SOMAARTH Data management software system: The unique feature of SOMAARTH surveillance is tracking of built environment and physical environmental parameters of a large geographically defined area besides individual-level demographic monitoring of the households. SOMAARTH has

Page 40 of 49

developed software systems to manage its demographic, development and environmental data streams, maintaining a consistent records of significant demographic events in the population, generate registration books, GIS maps that the fieldworkers can use, and compute basic demographic rates, such as birth, mortality, and total fertility. Further this provides opportunities to monitor developmental parameters such as change in land use, water bodies, building structurs, road density, retail desities, air quality level etc.

SOMAARTH DDESS captures different data streams which need to be matched and integrated into a single database for conducting analysis. In addition, in some cases data must be transformed into the preferred electronic format for a variety of statistical packages. Sometimes, data obtained from the field must be cleaned and debugged for input and measurement errors, and reformatted. For handling these issues SOMAARTH DDESS software (SOMAARTH-I) © developed by INCLEN. The technology used for the software is functional on web through PHP and on android platform through MySQL. It incorporates in itself capacity for handling huge datasets through MySQL with inbuilt quality assurance systems and trouble shoot mechanisms. The key features of the software includes: handling of population and census data, data sequencing through unique identification, data layering, project database creation. Software includes modules on registration of land parcels, user management, survey, quality assurance, query building, reporting (including tabular and graphical), cohort and multiple project management. Integrated web and android based data collection capabilities has made SOMAARTH-I software a robust package for handling data collection, storage, management and analysis for large volumes of longitudinal data sets.

Demographic events such as births, deaths, migrations, affect not only demographic rates, for instance, but also relationships within and between households. SOMAARTH DDESS also maintain socioeconomic data such as on marriage, family relationships, and economic conditions, owing to the strong correlation between health and socioeconomic status. As event histories grow, records of new events must be logically consistent with those of events in the past.

Data management starts with the indepth knowledge about the Residential and Non-residential instruments for building the data accuracy. Understanding of instruments is very much important before checking data. There are total 48 questions in the residential form and 14 questions in the non-residential form. Both the tools are available in English and Hindi format.

Steps in data management

• Step 1-Data Retrieval

- Step 2-Data Processing
- Step 3-Output of data

Step1-Data Retrieval: Question wise data are obtained in comma separated values (CSV) files for non-residential units, household (questions related to household information, e.g.-Type of house) and individual level (questions related to personal level, e.g.-Details of household members). The files which are usually retrieved from the standard SOMAARTH census database are shown in Box ?

Box 6.1 Data retrieval

- classification_non_residential: It contains data of Q5. Classification of non-residential buildings from NR form.
- dwellingunit: It contains data of Q1. Identification of dwelling unit of both the residential and non-residential forms.
- household: It contains data of Q2. Dwelling unit basic information of both the residential and non-residential forms.
- household_domestic_animal: It contains data of Q8. Details of the domestic animals in dwelling unit in the NR form and Q21.
- Details of the domestic animals owned by the household in the R form.
- household_furnishing: It contains data of Q22 Furnishing of the residential form.
- household_furniture: It contains data of Q22 Furniture of the residential form.
- household_health_facility: It contains data of Q18. Health facilities used by the household of the residential form.
- household_house_details: It contains data of Q10. Details of the dwelling unit and Q11. Construction of the dwelling unit of the residential form and Q6 and Q7 from non-residential form.
- household_income: It contains data of Q15. Monthly/annual income of the members of household of the residential form.
- household_land_activity: It contains data of Q13 (B), Q13(C), Q13 (D), Q13 (E), Q13 (F) of household of the residential form.
- household_land_ownership: It contains data of Q13 (A). Agricultural land of household of the residential form.
- household_land_sell_purchase: It contains data of Q12. Ownership of the land of household of the residential form.
- household_modern_equipments: It contains data of Q22. Details of Modern equipments in the household of the residential form.
- household_other_details: It contains data of Q9-Q14 of non-residential form and Q17, Q19, Q20, Q23-Q48 of the residential form.
- household_owner: Name of owner of residential and non-residential unit.
- household_person: It contains data of Q5A. Age details of individual along with ids in the household of the residential form.
- household_person_alcohol: It contains data of Q8. Alcohol consumption details of individual in the household of the residential form.
- household_person_chronic_disease: It contains data of Q9 (H). Chronic disease of individual in the household of the residential form.
- temp_household_person_details: It contains data of Q5A. Details of individual in the household of the residential form.
- household_person_disability: It contains data of Q7. Details of mental/physical disability of individual in the household of the residential form.
- household_person_substance: It contains data of Q8. Substance consumption details of individual in the household of the residential form.
- household_person_tobacco: It contains data of Q8. Tobacco consumption details of individual in the household of the residential form.
- household_social_status: It contains data of Q16. Social status in the household of the residential form.
- household_vehicle: It contains data of Q22. Details of the vehicles in the household of the residential form.
- household_visitors: It contains data of Q5D. Details of the visitors in the household for less than 6 months of the residential form

These original data are kept in a separate folder. In a different folder these CSV files are transferred to excel files for further working. Codebook was prepared for both residential and non-residential tool separately in excel.

Table 6.1: Example showing codebook format of residential tool					
Questio	Question	Туре	Codes	Codes	System Code
n	Descriptio			descripti	
	n			on	
Questio	Descriptio	Numeri	Codes given	What the	Codes developed in
n	n of that	c or	for response	codes	backend in the system
number	particular	string	in particular	stands for	against a particular
S	question		the tool		question. Sometimes it may
mention					different from the codes
ed in					mentioned in hard copy.
tool					These codes will be
					provided by IT team.
Q1	Block ID	Numeri	1	Hathin	1
		с	2	Palwal	2
			3	Hodal	3

Data processing:

Data cleaning: The data input process oftentimes introduces typos, miscodes, and errors into the data. The following steps were followed in data cleaning.

- Duplicative information occupies storage space, and in some cases give rise to data quality issues. Therefore datasets must thoroughly checked to identify duplicate values and duplicate values should be removed.
- Missing information in data should be treated carefully. Missing information should be checked question wise and then verified with hardcopies with the help of data manager handling hardcopy data archives.
- Inconsistencies were checked at two levels (a) Within variable (e.g.-if for a particular question codes are mentioned as 1,2,3 & 4 but in data additional code 5 is coming in responses) (b) Between variables (e.g.- If for a particular individual gender is coming as male and occupation is coming as housewife). In both the levels verification should be done by checking hardcopies in coordination of data manager. After verification datasets should be rectified.
- All households have defined members at any given point in time (rules unambiguously exclude nonmembers);
- All households have a single head at a given point in time, and members relate to one another and to the head in definable ways;
- Members have names, dates of birth, and other characteristics that do not change over time;
- Events can occur to members, such as death, birth, in- and out-migration, and marital-status change (attempts to enter event data on nonmembers are rejected at the point of data entry);

- Events change household membership and relationships according to fixed rules; and
- Episodes (such as pregnancies, conjugal relationships, or residencies) are associated with individuals at risk (that is, active members) and must follow simple logical rules.

Data Merging: SOMAARTH 1 surveillance software provides question wise datasets. Merging of all the question and creation of single table should be done at household or non residential unit level and individual level separately. This could be done in any data management software such as STATA 12 or SPSS and a coding manual should be written that matched all variable names with variable labels and codes. Final data set could be made available in excel and stata format as per the needs of the researcher. **Output of data:** Basic frequency tables should be prepared using residential and non-residential unit level as well as parcel level datasets.

The format followed for preparation tables-

Box1 : Format for basic frequency tables										
S1.	Village	Villag	Resid	Non -	Residential			Non - Residential		
No.	Name	e code	ential	Residenti al	Loc ked	Ope n	Rui ned	Loc ked	Ope n	Rui ned
1	ANDHR OLA	AND	566	253	32	532	2	15	227	11

Preparation of tables: Basic descriptive tables need to be prepared using standard data format of National Family Health Survey, District Level Health Survey report.

Factsheet preparation: Dissemination of data form the key function of conducting surveillance activity. For this factsheet indicating key variables of the surveillance site should be prepared for all the villages / wards under the surveillance site. It gives the information on the geographic features, socio-economic and demographic features, health and substance use and household features. These indicators can be are then matched with sustainable development goals (SDGs).

		Box 6.2: SOMAARTH DDESS Key Indicator Template
Geog	raph	ic features
	≻	Total Area (in km ²)
	>	Built-un area (in km ²)
	Ś	Total land narcels
	6	Posidential (%)
		Non Desidential (%)
	-	
	>	
	>	Others (%)
•	Mix	ed (%)
•	Lan	d parcel density (km ²)
•	Ave	rage size of constructed land parcels (mt ²)
•	Pop	ulation density (km ²)
•	Stru	ictural unit density (km ²)
•	Roa	d density (length ner km ²)
Socio		nomic and demographic features
buch	> <u>CCOI</u>	% of female
	6	Sex-ratio (females per 1000 males)
		% of child_nonulation (0.6 years)
		Child son notio (ginls non 1000 hours agod 0, 6 years)
		% of old one nonvlotion (60 + views)
		Child women wetic (25 children ner 1 000 wemen aged 15 40)
		Child dependency ratio (nonvlation area 0, 14 to 15, 64 years)
		Clinic dependency ratio (population ages 0–14 to 15–04 years)
	-	Old-age dependency ratio (population ages 65+ to 15-64 years)
	-	% of mate interates (0+ years, who cannot read and write)
	~	% of remain interates (6+ years, who cannot read and write)
	~	% of school-going age children(6–8 years) who are not attending school
	~	% of population belonging to Hindu
	~	% of population belonging to Niuslim
	~	% of population belonging to SC (dominant community: Jatav-58.0%)
	~	% of population belonging to OBC (dominant community: Meo-56.8%)
	~	% of population belonging to General (dominant community: Jaat-72.9%)
	~	% of unemployed (18–60 years)
	~	% of population (18–60 years) involved in farming
	>	% of population (18–60 years) involved in Govt./private jobs
	>	% of population (18-60 years) involved in daily labor/NREGA/other contract works
	>	% of single/unmarried males (20–29 years)
	>	% of single/unmarried females (20–29 years)
Heal	th &	substance use (self reported)
•	% s	uffering with any chronic health problem (15+ years)
•	% s	uffering with any chronic health problem (60+ years)
•	% v	vith any (mental/physical) disability (all-ages)
•	% 0	f males, who consume tobacco (15+ years)
•	% 0	f males, who consume alcohol (15+ years)
•	% 0	f males, who consume alcohol or tobacco with any chronic health problem (15+ years)
Hous	eholo	l features
•	Ave	rage household size (number of persons per-household)
•	% (f female headed-households
•	%	f houses with complete pucca structure (roof, wall and floor)
	%	f households with electricity connections
-	%	f households using primary fuel for cooking – Riomass & Agricultural residue
	0/2 0	f households using secondary fuel for cooking _ I DC
•	70 C	f households using unimproved water for drinking & cooking
-	70 C	f households with toilet facility
•	70 C	n nousenonus with tollet lachty
•	% C	n nousenous fonow unsafe sond waste disposal practices
•	∽% C	i nousenoids without any land (ownership)

Socio economic Index: The SES tool used for SOMAARTH considers household assets and housing characteristics. In NFHS-3, 33 indicators were used for developing wealth index. Out of these 33 indicators, the SOMAARTH SES tool captures information on 31 indicators. Scores are assigned to these 31 indicators on the basis of SLI, NFHS-2: 4 each for a pucca floor, pucca roof, pucca wall, a car, a tractor, 3 each for a computer, refrigerator, motor cycle or scooter, 2 each for a telephone, bicycle, radio or transistor, electric fan, black and white TV and 1 each for a mattress, chair, table, cot or bed, watch. This allows grouping the households according to their SLI positions using cutoffs developed on NFHS data.

Chapter 7: Ethics Considerations

A demographic surveillance site is a geographically defined population with continuous monitoring and regular production of data and reports on basic demographic event such as birth, death and migrations. As per guidelines given by the Indian Council of Medical Research (2017) setting up such sites in India need not be subjected to prior review and approval by an Ethical Committee if prior approval from competent state/ national authorities and from the community leadership has been taken.

Data Safety and confidentiality:

SOMAARTH database are managed at two different levels i.e., geospatial database (land parcel) and census enumeration database. Every land parcel and their residential and non residential units falling within the surveillance area are identified and allotted a composite Unique Identification Number having 30-digits which includes country code, state code, district code, block code, village code, sector number, land parcel number, unit number. Each individual registered within the surveillance site are allotted a 9-digit unique identification number which is randomly generated by the surveillance software (SOMAARTH 1). Location identification and individual UIDs are kept seprate to each other in order to safe guard the personal locational details of the individuals registered within the site. Both the UIDs work independently and when both are combined personal details of the population are blinded. As a practice hardcopies of data are kept locked in keys and softcopies are kept as encrypted files in computer system. In order to protect the confidentiality of personal information the individual-level data collected are not be shared. There is a restricted access to the geospatial data. In the special cases where geospatial data is required, it is mandatory to conceal the personal details of the individuals adopting the reverse geocoding system. The identification details of subjects is concealed through software generated random number UID system.

Box 7.1: SOMAARTH DDESS Palwal has following agreements /partnerships with the different departments of Government of Haryana, India.

- MOU signed with Govt. of Haryana 2012
- Coordination mechanism at District and State level
- SOMAARTH Team is part of District Development Review Board
 - **Active coordination with Departments**
 - **District Health Services**
 - **District Health and Family Welfare Society**
 - **District Education Office**
 - **District Information and Public Relations Office**
 - **District Agriculture Office**

Chapter 8: SOMAARTH DDESS Administration

The policies on employment, staff benefits and remunerations, training and staff development are dealt in details in the Manual of Standard Operating Procedure of IEO and the same is applicable to the DDESS as well.

Outlien of Administrative/Core governance Structure: Field Operation



Figure 6.1: Core governance structure

Recruitment& training:

All recruitments shall be done through HR Department at IEO. With due requisition and sanction from competent authority, the HR Dept initiates the process of recruitment. Positions are advertised in newspapers, job oriented professional websites and also loaded on INCLEN/ SOMAARTH website. Depending on situations/posts, advertisements are also sent to Sarpanch of DDESS villages. In addition, application forms shall also be made available in the villages at Sarpanch house for candidates on expiry of last day, all received CVs/ Bio-Data/ Applications are short listed according to advertised age, experience and academic qualifications. For recruitment of technical and staff like field staff and field supervisor there shall be a written screening test. Short listed candidates are to be issued with admit cards to appear in written test. On the basis of written test a merit list is prepared to call for interview

and field test. On the basis of written test, interview and field test (if necessary) a final merit list is prepared. As per requirement of the activities top listed candidates are asked to join. Selection and recruitment processes are unbiased and do not accommodate any recommendations/ requests.



Figure 3.2: Surveillance Activity Operational Diagram