

2.3 Methodology

2.3.1 Survey Preparation and Approach

After a series of preparatory and consultative meetings, The Gambia's Scan-ICT baseline study took off with the commencement of desk research in April 2006. This activity involved members of Scan-ICT technical committee representing the various stakeholders. The desk research lasted about 4 weeks during which secondary data was gathered and where not available the possible source was detailed.

This activity continued through the design stage of the survey when the necessary instruments for the subsequent data collection activities, intended to fill the gaps, were drafted. A survey methodological targeting both institutional and household was agreed.

The institutional survey covered those key institutions and employees that Scan-ICT focused on and, household survey targeted households and individuals access to and exploitation of ICT.

2.3.2 Sampling

A preliminary review of the pattern of infrastructural development and expansion in the Gambia suggest there is some kind of relationship between areas in terms of access to certain ICT infrastructure and facilities. As a result, stratification was done accordingly using both the population data and preliminary figures on electricity availability in households from the 2003 Population and Housing Census. Four strata were created:

- Stratum 1 comprises of Banjul and Kanifing;
- Stratum 2 combined Brikama urban and Kombo North;
- Stratum 3 constitute of all Other Urban and
- Stratum 4 consisted of all the Rural areas.

The two-stage sampling involved the selection of 2003 updated Enumeration Areas (EAs) as the primary (first stage) sampling units (PSU) and, the final selection of households as the secondary (second stage) sampling units, after the listing exercise of

households in an EA. Seventy-six enumeration areas were selected as the primary sampling units and twenty households selected in each EA. Final sample of households obtained is 1504 households.

Box 1: Scope of Survey

The survey collected information from 201 institutions and 1504 households through out the country to meet the stated objectives of the baseline study. The institutions were grouped into six major categories: Government; Parastatal; Non Governmental Organizations (NGO); ICT sector; Private Business and Health sector.

There are two sources of data: primary data sources and secondary data sources. Primary data sources come from direct collection of data from institutions (201) and households (1504) persons (1428) representing individuals in households, employees(440), students(477) and teachers (227) in the education sector, that is a total of 4277 sample units. The distribution of the sample is presented in the following

Table 3: Distribution of samples per Local Government Area

Local Government Area	Government	Health sector	Education sector	ICT sector	Business sector	Households
Banjul	8	1	1	2	19	80
Kanifing	6	1	36	9	29	395
Brikama	3	3	2	1	0	399
Mansakonko	3	1	14	0	0	80
Kerewan	6	2	26	0	0	193
Kuntaur	2	0	5	0	0	80
Janjangbureh	2	1	13	0	0	118
Basse	2	1	0	0	0	159
The Gambia	32	10	97	12	48	1504

Source: SCAN –ICT Baseline Survey 2006, The Gambia

This survey had the double preoccupation of collecting data not only concerning institutions but also that concerning households and individuals. Secondary data sources are existing statistics from international organizations, public and private administrations, etc.

2.3.3 Questionnaire and training of personnel

There are nine types of questionnaires (see annex) used: the Government/Parastatals/NGO questionnaire, Private Business questionnaire, ICT sector questionnaire, Employees' questionnaire, Household questionnaire and Individual questionnaire, educational institutions, students and teachers questionnaire. After the instruments were discussed and finalized by the Technical Committee through series of meetings, the baseline study moved into survey stage. It started with training of data collectors for duration of two days after which data collectors were sent out into the field to administer the questionnaires. The objective of the training was to introduce the data collectors to the instruments and discuss the technical questions. It also helped the data collectors to know how to administer instruments.

2.3.4 Deployment of enumerators in the field

Teams of data collectors were composed and assigned to collect information from sampled institutions and households throughout the country. Each team of data collectors consisted of a supervisor and enumerators and were assigned number of selected enumeration areas in a geographical settings to cover. A vehicle was provided to each team.

2.3.5 Estimating Model for household and individual survey

After a through review of our sampling design and, the need for an appropriate estimator among competing estimators, we at last found a model estimator for our producing of reliable national and domain estimates which also enables us to compute variances for our proposed estimates.

The model estimator so obtained is based on our Stratified two-stage sampling design and, is as follows:

$$\hat{P}_{st} = \frac{\sum_{h=1}^L N_h M_h p_h}{\sum_{h=1}^L N_h M_h} = \sum_{h=1}^L W_h p_h$$

Where, $W_h = N_h M_h / \sum_{h=1}^L N_h M_h$ is the weight or relative size of stratum h in terms of second-stage units and \hat{p}_h is the estimated proportion in stratum h.

So, an unbiased estimator of the variance of the estimated proportion is:

$$\hat{V}(\hat{P}_{st}) = \sum_{h=1}^L W_h^2 \left[\frac{1-f_{1h}}{n_h} s_{1h}^2 + \frac{f_{1h}(1-f_{2h})}{n_h m_h} s_{2h}^2 \right]$$

Where, $f_{1h} = n_h / N_h$, $f_{2h} = m_h / M_h$ are the respective sampling fractions;

and, $s_{1h}^2 = \frac{\sum_{i=1}^n (p_i - \bar{p})^2}{n-1}$, $s_{2h}^2 = \frac{m}{n(m-1)} \sum_{i=1}^n p_i q_i$ are the respective sample variances;

and, $p_i = y_{ij} / m$, $\bar{p} = \sum_{i=1}^n p_i / n$

Our Standard Error and Confidence Interval (CI) for the estimate respectively, will be:

$StdErr(\hat{P}_{st}) = \sqrt{\hat{V}(\hat{P}_{st})}$, and, $CI = \hat{P}_{st} \pm StdErr(\hat{P}_{st}) \times t_{df, \alpha/2}$ (d.f.=no. of EAs minus no. of strata in the design)

Now, let's define the notations as follows:

- $h = 1, 2, \dots, L$ is the stratum number, with a total of L strata,
- $i = 1, 2, \dots, n_h$ is the EA number within stratum h, with a total of n_h EAs,
- $j = 1, 2, \dots, m_{hi}$ is the number of units within EA i of stratum h, with a total of m_{hi} units
- $n = \sum_{h=1}^L \sum_{i=1}^{n_h} m_{hi}$ is the total number of observations in the sample

- y_{hij} is the observed value of the analysis variable for observation j in EA i of stratum h, for both the values of numerical variables and the values of indicator variables for levels of categorical variables.

2.3.6 Collection of data

The data collection activity for both institutions and households lasted for twenty-one days during the months of July and August 2006.

2.3.7 Data Processing and Cleaning

After the instruments were finalized, ten data entry screens were designed using CSPro Version 3.1 and commissioned. These covered institutional, household and individual questionnaires - employees and household individuals. The screens were designed

Following the completion of field activity, the batches of completed questionnaires were scrutinized in the office by a team of coders, followed by editing and coding exercise. Having edited and coded the batches, the data entry activity was carried out by a team of data entry operators over a period of two weeks.

The entered data was then checked for errors or wrong entries which involved the generation of frequency tables. Editing was done where necessary and the cleaned data was then imported into SPSS for analysis. At this stage the data-set were put in a form according to the thematic areas of Scan-ICT. Statistical tables were produced from the datasets forming the basis of analysis, estimates and the production of the report.