



A household data collection system with a continuous multipurpose survey (INCAF)

**Report from a short term mission to the National Statistical Institute
of Mozambique, Maputo Mozambique**

6-26 May 2012

within the frame work of the

***AGREEMENT ON CONSULTING ON
INSTITUTIONAL CAPACITY BUILDING,
ECONOMIC STATISTICS AND RELATED AREAS***

between

INE and Scanstat

Lars Lundgren



INSTITUTO NACIONAL DE ESTATÍSTICA

Contents

SUMMARY	3
Resumo em Português.....	3
CONCEPTS AND ADVANTAGES	4
A National Household Data Collection System.....	4
Traditional approach	5
New “light” approach.....	5
POSSIBLE INDICATORS IN INCAF RELEVANT FOR PARP AND THE FIVE YEARS PLAN.....	5
SAMPLING DESIGN AND FIELD ORGANISATION	13
IT SOLUTION	14
Database design.....	14
Data entry	15
Data communication and storing	15
Editing and coding	15
Data manipulation	16
Tabulation	16
QUESTIONNAIRE DESIGN	16
WELFARE/POVERTY MONITORING	18
Poverty line	18
Poverty headcount and poverty rates	18
Engel’s Law	18
DISSEMINATION.....	19
RISKS AND SOLUTIONS.....	19

Summary

There is a pronounced need for more timely socio-economic statistics. Statistics about the labour market only every five year is not relevant. Macroeconomic and poverty reduction policies also need more current consumption data. INE is prepared to solve the problem with a continuous multi-purpose survey (INCAF) giving quarterly labour force data according to ILO recommendations. The quality of consumption data will be enhanced compared to the latest household budget survey (HBS), particularly for poverty classifications. Most other household related data can be collected through rotating modules.

INE is well on track to start the INCAF 1 July 2012 after having conducting a successful pilot. The recommendations are fine tuning implementation for more efficient sampling design and distribution of questions/modules over time to create a more even workload for the field staff and the respondents.

To measure more qualitative indicators and the citizens' satisfaction a Service Performance and Quality of Life module is presented.

Data are going to be collected using PC-tablets, which will avoid a common bottleneck, but it is important that all management of data is robust and efficient to allow a continuous flow of data and safe linking of different types of data.

Resumo em Português

Há uma necessidade evidente para estatísticas socioeconómicas mais oportunas. A produção de estatísticas sobre o mercado de trabalho apenas a cada cinco anos não é relevante. Políticas macroeconómicas e de redução da pobreza também precisam de dados de consumo mais actuais. INE está preparado de resolver o problema com uma pesquisa multiuso contínuo (INCAF) apresentando dados trimestrais da força de trabalho de acordo com recomendações da OIT. A qualidade dos dados de consumo será reforçada em comparação com a última Inquérito sobre Orçamento Familiar (IOF), particularmente para as classificações de pobreza. A maioria dos outros dados dos agregados familiares pode ser recolhida através de módulos rotativos.

INE está no bom caminho para iniciar o INCAF 1 de Julho de 2012 depois de ter realizando um piloto bem-sucedido. As recomendações são de ligeiramente ajustar a implementação com um design mais eficiente de amostragem e a distribuição das perguntas / modules ao longo do tempo para criar uma carga de trabalho mais equilibrada para o pessoal de campo e os respondentes.

Para medir os indicadores mais qualitativos e a satisfação dos cidadãos um módulo Desempenho dos Serviços e a Qualidade da Vida (DSQV) é apresentado.

Dados vão ser recolhidos usando *PC-tablets*, o que evitará um estrangulamentos comum, mas é importante que toda a gestão de dados é robusto e eficaz para permitir um fluxo contínuo de dados e uma ligação segura entre os diferentes tipos de dados.

(O relatório inteiro em Português pode-se encontrar no www.dst.dk/mozambique)

Concepts and advantages

The tradition in most developing countries has been to make household surveys as standalone surveys, sometime following a five year program. Those programs have often been interrupted. With a faster developing society and economy it is not very relevant to have a labour force survey once in five years time. Employment data, as a leading economic and social indicator, should be quarterly. Starting a new survey every year, with planning, piloting and training of new interviewers is both costly and time consuming. With ambitious data cleaning and analysis, the results have often come out late and with lack of relevant information, and funders are not eager to put in more money.

When doing ad hoc surveys it is tempting to go outside the core scope of the survey and include many other welfare measures, sometimes even forgetting the main scope (the latest DEMOGRAPHIC and Health Survey is not giving any data on the population under 15 years old). Another problem is that the ad hoc or programmed survey can be affected by temporary events, like natural disasters, not being able to capture the longer trends.

The concept of a continuous survey or data collection system is the opposite. It should be a well-trimmed machine, all the time collecting data as they are needed, but spread over time as much as possible to give an even work load for the machine as well as the respondents. Very few data are needed quarterly and some data like time use and food intake change slowly and can be measured with significant changes perhaps every ten year. The users don't need to jump on a crowded train, as a new one soon will leave.

It would be good to talk about a household data collection system rather than a continuous multipurpose survey, to avoid the thinking on a survey as a readymade tool, instead of an adjustable tool.

The main advantages with a permanent data collection system are:

- Faster dissemination of statistics by updating annual standard publications and quarterly trends
- Increased user awareness by regular dissemination
- Synergies in output with comparability of the total database, not only for the background variables
- More flexibility to include upcoming needs
- Stable field organisation and standardisation of all other processes, supporting higher quality
- Lower costs (less planning and training)
- A regular budget can easier get sustainable funding.

A National Household Data Collection System

To facilitate standardization, cooperation, coordination and efficiency, the survey can be extended to National Household Data Collection System. The system should be viewed as the obvious mechanism for high-quality data collection and processing of data from household samples. It should also be viewed as a core part of the overall national statistical system and include, engage and train the users for efficient use of the results.

The system should be process and output oriented, meaning that the organisation and all parts of the system should support an efficient and high quality process from collection, through data processing to a database with easy access. The output should be relevant, reliable and timely to facilitate evidence based decision making.

The core of the survey should be fairly slimmed to the most necessary data that need to be monitored quarterly or annually. Other data should be collected through add-on modules rotating over time and be collected from sub-samples when a smaller sample is adequate. The core questionnaire should not take more than around one hour to complete. Each add-on module should be limited to around 30 minutes.

Traditional approach

Common in many countries, is to have a have a continuous Household Budget Survey with a smaller sample and measure all consumption with one month (or two-four weeks) diary keeping, supported by weekly visits. With many visits it is easy to add modules without over-burden the households.

New “light” approach

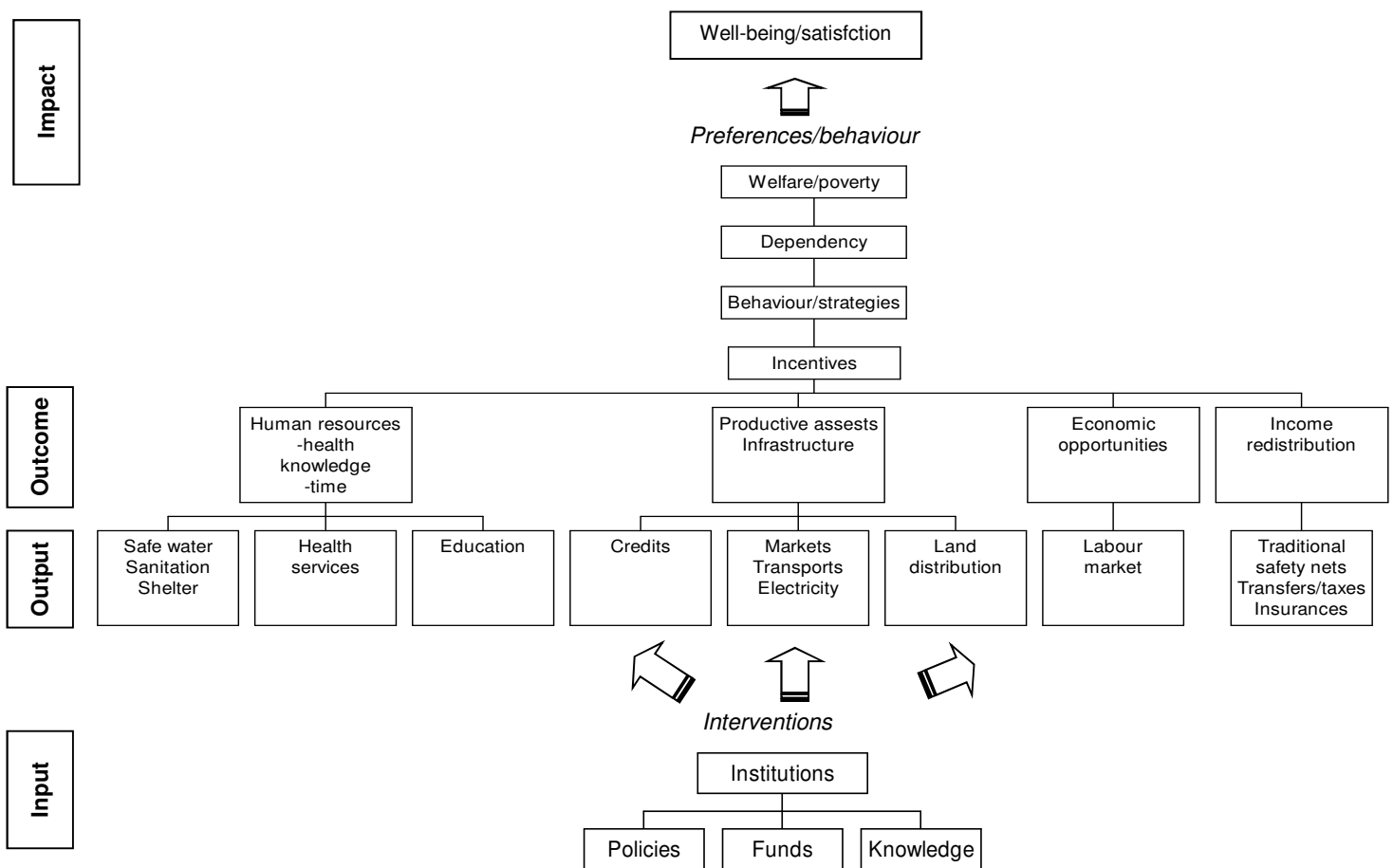
It is difficult to remember the purchases a longer time. In fact, a household buying many things at one time will probably have difficulties to remember all bought items in details already when they have been put into the shelves and fridge. That is why a daily diary recording is necessary for those buying a lot and probably can read and write, while poorer households can remember the little they bought. In general it is not possible to ask for aggregates of consumption, e.g. vegetables. We don't buy “vegetables”, but tomatoes, cucumber, etc. and to ask such a question we have to summarise all bought vegetable, which is more difficult than give an answer for each vegetable. There is, however, one exception and that is the lump sums of what is paid at each purchase (food, beverages, cleaning and washing items and other consumables). The households should be able to remember those lump sums for the last seven days. Visiting the households four times (once in each quarter), it is possible to include seasonal differences for each household (for better poverty classification). The panel approach will give more accurate change estimates particularly for labour force data.

Possible indicators in INCAF relevant for PARP and the Five Years Plan

Poverty is a multidimensional phenomenon, and combating poverty goes well beyond a simple discussion of the underlying characteristics of absolute poverty. Rather, it is an issue that needs to be addressed from a broader perspective, reflecting the fact that “individuals, families and communities lack the capacity or the opportunity to gain access to minimum living conditions according to the basic standards of society.” (PARP)

It is common and useful to have indicators for the mean-end logic for monitoring interventions and cause-effect indicators for deeper analysis. In the *input* → *output* → *outcome* → *impact* chain, *input* → *output* represents the production/supply perspective, while *outcome* → *impact* represents the user/demand side. To minimize the burden on a continuous survey, it is optimal to focus the household survey on *outcome* → *impact* and leave *input* → *output* to accountings and other administrative data.

To humanize those terms they can be transferred to an *availability (output)* → *incentives* → *access (welfare)* → *behaviour/preferences* → *utilization* → *satisfaction/vulnerability (well-being)* chain. Access to assets and services are the most common measures in welfare surveys, not considering preferences and the actual utilization of the resources. Quality aspects are not easy to measure objectively and therefore often not included. Previous figure tries to give a simplified overview of main aspects of welfare and well-being in a policy relevant mean-end (or cause-effect) logic. A more comprehensive map of well-being is found in annex 1. The map can be useful for selection of relevant indicators to include in the INCAF.



From welfare to well-being

During the last decade the interest has increased for stretching from welfare to well-being/satisfaction. Some statisticians argue that it is difficult to include people's preferences and measure satisfaction in objective ways. But, in fact, it is not that straight forward to measure access either. E.g. a new road may not be so accessible if the household has no car, if there are no busses or if it is too expensive to use it. The same with health care, which you not even need if you are healthy (and sometimes the quality of the care is at a level when it is better to stay away). Distance or time to nearest hospital may look very objective but may not be very relevant, if a lot of supplementary questions are not asked.

Many attempts have been made to measure well-being by making composite indices. The problem is to put weights on the different components. Commonly, experts are doing it and the results are that Scandinavian indices put Scandinavia in top, while Dutch indices put the Netherlands in top. Only the people "know" the weights, but only unconsciously. Asking how important health care is, most people will put on a high weight. And so they will do on education, irrespective of their own level, and most other things.

Another way to include preferences is to ask about total satisfaction of something, say the school, and then satisfaction with different aspects of the school, e.g. distance to school, materials, classrooms, food at school, teachers, teaching methods, homework, grading system, protection, personal treatment, etc., including quality aspects. The indicators can be grouped into components, e.g. access, teaching quality, safety and participation (before or after the data collection). All indicators are measured on a scale, preferably from 1 to 10. The analysis is made with a method called *Structural Equation Modelling with Latent Variables*, using the *PLS (Partial Least Squares)* technique.

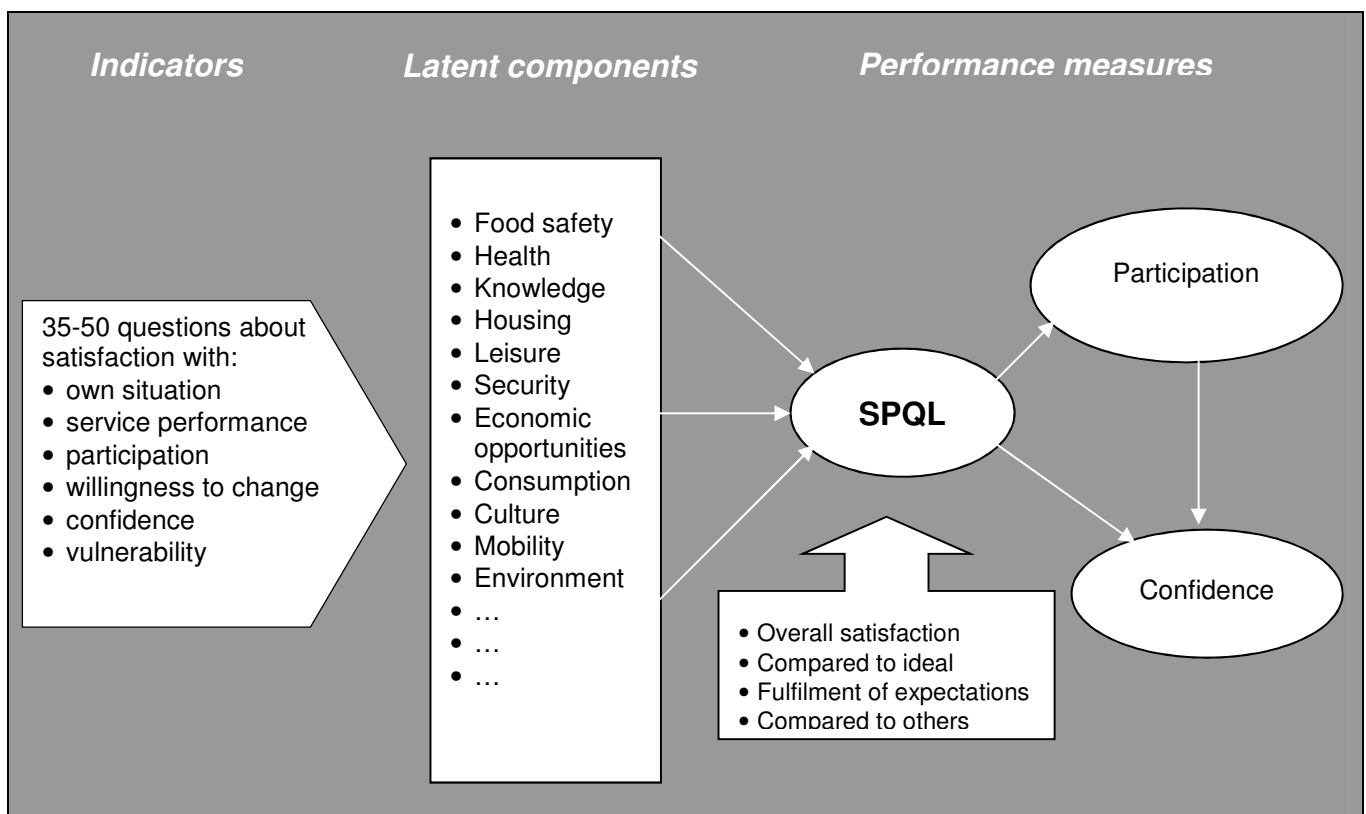
The result of the analysis gives

- For each component an *index* value between 0 and 100 indicating the satisfaction with this specific component
- For each component an *impact* value denoting the impact of the component on the overall satisfaction.
- The overall satisfaction with service performance and quality of life.

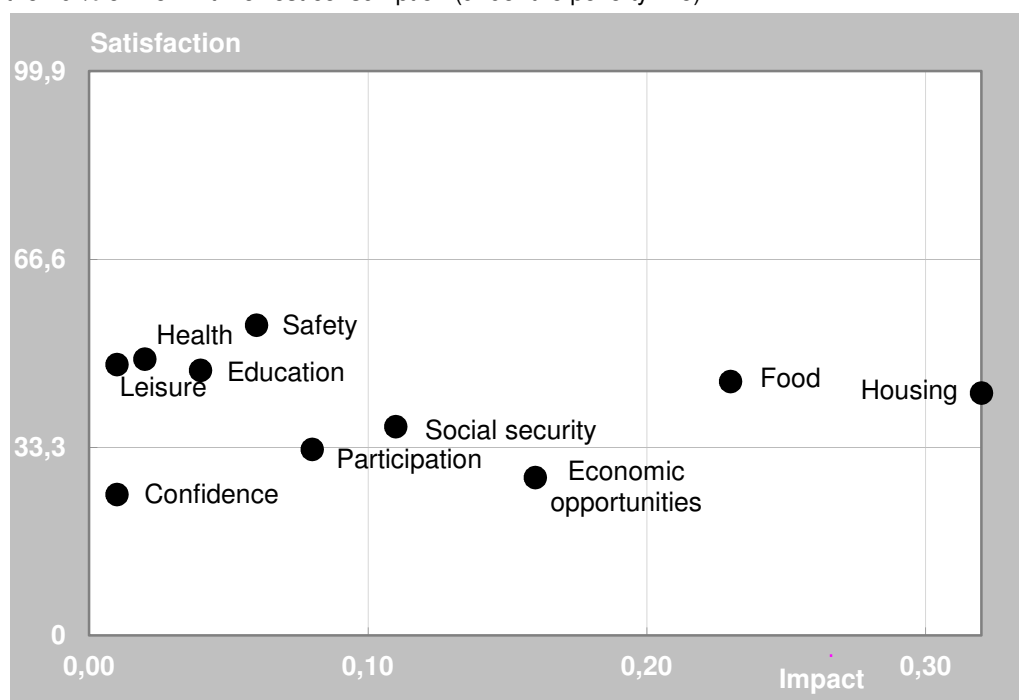
Thus, the analysis identifies the components that policy makers and managers should focus on for efficiently increasing the customers' or citizens' overall satisfaction. In the analysis the index values are attached to each respondent. Using background variables, it is possible to identify subgroups for further analysis.

This method is the state of the art to measure quality and service performance (SP) in the private sector in EU, US, Australia, Japan, and Singapore and in other countries. Statistics Sweden is regularly monitoring child care, health care, care of elderly, schools and citizens' satisfaction throughout Sweden.

This method can be zoomed in, e.g. to measure staff satisfaction or zoomed out to measure quality of life (QL) or combined to measure Service Performance and Quality of Life (SPQL) at the same time. It is also feasible for measuring children's well-being in their own perspective.



Satisfaction and impact on total satisfaction by well-being components in Jamaica, 2003, the 40 % of men with lowest consumption (under the poverty line)



Policy makers in Jamaica should focus on housing, food and economic opportunities, as those components have most impact on total satisfaction (for poor men).

It is also possible to group the indicators in other dimensions, e.g. access to services, quality of services, own situation

This type of survey has been carried out in following countries, as ad-on modules in multi-purpose surveys or household budget surveys supported by Statistics Sweden or stand-alone surveys supported by the Swedish Children Ombudsman.

Country/Group	SPQL
Children (9-16) in Sweden, 2002	88
Children (9-15) in Serbia 2004	84
Children (7-15) in Montenegro 2005	77
Children (8-18) in Sri Lanka 2004	74
Children (9-16) in Vietnam 2004	73
Children (8-17) in Peru	66
Children in Kosovo (9-15), 2003	60
Adults in Jamaica 2003	54
Adults in Oman 2000	46
Adults in Abu Dhabi 2008	44
Adults in Serbia, 2003	41
Adults in Lesotho 2003	40
Adults in St Petersburg, 2002	39
Adults in Kosovo, 2003	34

Annex 2 provides an example of a SPQL questionnaire.

In most ad hoc surveys like HBS, MICS and DHS it is common to collect as much information as possible. It is a onetime chance. With a continuous survey (or system) like INCAF, the data collection

should be spread as much as possible, to get a more even work load. With four visits per household and with rotating modules within the year and between years it is possible to collect almost any household information without a heavy response burden. Here is a tentative list generated from the PARP.

General objectives	Indicator	Level	Source	Periodicity ¹
Increase output and productivity in agriculture and fishery	1 Land	Access to Resources	SPQL +Core	5+A
	2 Irrigation	AR	SPQL +Core	5+A
	3 Credits	AR	SPQL	5
	4 Roads	AR	SPQL	5
	5 Markets	AR	SPQL	5
	6 Telephone	AR	SPQL	5
	7 Storage facilities	AR	SPQL	5
	8 Mineral fertilisers	AR	SPQL	5
	9 Organic fertilisers	AR	SPQL	5
	10 Improved seeds	AR	SPQL	5
	11 Insecticides	AR	SPQL	5
	12 Animal vaccines	AR	SPQL	5
	13 Animal drugs	AR	SPQL	5
	14 Extension services	AR	SPQL	5
	15 Producers' organisations	AR	SPQL	5
	16 Sold agricultural products	Behaviour	Core	A
	17 Agricultural vulnerability/food security	Impact	SPQL +Core	5+A

General objectives	Indicator	Level	Source	Periodicity
Foster human and social development	18 Birth registration	AR	SPQL	5
	19 Maternal and neonatal care	AR	SPQL	5
	20 Health services	AR	SPQL	5
	21 Clean water	AR	SPQL	5
	22 Water treatment and storage	Behaviour	Community	3
	23 Food distribution	AR	SPQL	5
	24 School nutritional support program	AR	SPQL	5
	25 Sanitation	AR	SPQL	5
	26 Medicine	AR	SPQL	5
	27 Birth attendance	AR/Behaviour	SPQL	5
	28 Vaccination	AR/Behaviour	SPQL	5
	29 Insecticide treated bed nets	AR/Behaviour	SPQL	5
	30 Hygiene practice	Behaviour	Community	5
	31 Utilisation of health services	Behaviour	DHS	5
	32 Maternal nutrition	Behaviour	DHS	5
	33 Infant and child feeding practices	Behaviour	DHS	5
	34 Iron and vitamin A	AR/Behaviour	DHS	5

¹ Q: Quarterly, A: Annual, 3: Each third year, 5: Each fifth year, 10: Each tenth year

35 Infant mortality rate	Impact	DHS	5
36 Under five mortality rate	Impact	DHS	5
37 Maternal mortality rate	Impact	DHS	5
38 Malnutrition (weight for height/ age)	Impact	Core	A
39 Incidence of diarrheal diseases	Impact	Core	Q
40 Malaria prevalence	Impact	Core	Q
41 HIV/AIDS incidence	Impact	DHS	5
42 Demographic growth	Impact	Core	A
43 Body Mass Index (BMI)	Impact	(Core)	A
44 Degree of physical activities	Behaviour	Core	A
45 Literacy rates	Impact	Core	A
46 E-literacy rate	Impact	Core	A
47 Access to education	AR	SPQL	3
48 Education quality	AR	SPQL	3
49 School enrolment	AR/Behaviour	Core	A
50 Drop-out by cause	AR/Behaviour	EDUC	3
51 Quality of housing	Impact	Core	A
52 Ownership of durable goods	Impact	Core	A
53 Pollution	Impact	SPQL	3
54 Safety from landmines	Impact	SPQL	3
55 Social protection	AR/Impact	SPQL	3
56 Income and consumption inequality/concentration	AR/Behaviour/Impact	Core	A
57 Drought, flooding, death, illness	Shocks	Community	?
58 Quality of life	Impact	SPQL	3

General objectives	Indicator	Level	Source	Periodicity
Promote employment	59 Vocational training	AR	Core + SPQL	A+3
	60 Labour market information	AR	Core + LM	Q+A+3
	61 Access to electricity	AR	Core + SPQL	A+3
	62 Informal sector development	Impact	Core + LF	A+3
	63 Employment by sector and quarter	Impact	Core	Q+A
	64 Working conditions	Impact	SPQL	3
Good governance	65 Business climate (laws and enforcement)	AR	SPQL	3
	66 Satisfaction/confidence in decentralised governance	Impact	SPQL	3
	67 Reducing waiting times	Impact	SPQL	3
	68 Demand for services	Behaviour	SPQL	3
	69 Tax payments	Behaviour	SPQL	3

Macroeconomic and fiscal stability	70 Real GDP growth rate	Impact	Core	Q+A
	71 Maintain consumers' purchasing power (Consumer Price Index)	AR/Impact	Core	(A)
	72 Local prices	AR/Impact	Community	Q+A
	73 Remittances from abroad	Behaviour		Q
<i>Mainstreamed issues:</i> Poverty reduction Gender equality	74 Share of population living in poverty	Impact	Core	A
	75 Food share of total budget	Impact	Core	A
Human and child rights, democracy	76 Ratio boys and girls by school level	Impact	Core	A
	77 People's full economic activities	Impact	Time Use	10
	78 Women's decision making (family planning, household economy, etc.)	Behaviour	?	10
Decentralisation	79 Degree of participation in local activities by sex	Impact	?	10
	80 Risk for violence (by sex), the effects of criminality and corruption perception	Impact	SPQL	5
Environment protection	81 Children's well-being	Impact	SPQL	5
	82 Child labour	Behaviour	Module	10
	83 Central and provincial service performance	AR	SPQL	5
	84 ICT in services	AR	Community	5
	85 Early warning system	AR	Community	5
			Community	5

The Government will focus on improving data quality related to the quarterly national accounts. (the Five Year Plan). INCAF can support quarterly national accounts as well as regional accounts.

Gender is a cross-cutting issue and has to be integrated in other policies. In an economy mainly based on subsistence farming and informal businesses it is important to measure all economic activities, whether they are paid or not and whether they are made by men or women. This can be done by a fairly simple time use module.

Children (up to 18 years) make up almost 50 % of the population, but are only marginally visible in official statistics. They are usually only visible as breakdowns by age of general tables. Almost nothing is known about their well-being. A child well-being module can easily give such information.

The described indicators and modules (questionnaires) should only be seen as examples and possibilities. They have to be fine-tuned to better facilitate national and regional needs. Such fine tuning should be done in a transparent process with main stakeholders.

Indicator/variable	Visit 1	Visit 2	Visit 3	Visit 4
Household characteristics and education	X			
Labour force (non-structural)	X	X	X	X

Labour force (structural)		X		
Housing		X		
Non-durable expenditure last week	X	X	X	X
Consumption of own produced food yesterday	X	X	X	X
Semi-durable expenditure last 3 months	X	X	X	X
Durable goods past 12 months				X
Remittances from Mozambique	X	X	X	X
Remittances from Abroad	X	X	X	X
Possession of durable goods				X
Consumer confidence	X	X	X	X
YEAR 1 rotating modules				
Tourism last 3 months	X	X	X	X
Child labour (and YEAR 11)			X	
(Service Performance and Quality of Life) (and YEAR 6 and 11) (sub-sample?)				(X)
YEAR 2 rotating modules (and YEAR 5 and 8)				
Detailed consumption (with diary and 4 extra visits per household), income	X	X	X	X
YEAR 3-4 and 6-7 rotating modules				
Food intake (each 10 year) (sub-sample)			X	
Time use (each 10 year) (sub-sample)	X	X	X	X
Informal sector (each third year)				X
Education (each third year)				X
ALL TIME				
Some space for urgent needs	X	X	X	X

Agricultural production and health care could also be integrated, but they are well covered by existing surveys.

Sampling design and field organisation

The sample for a multi-purpose survey can be more difficult to design as it has to provide reliable estimates for many different indicators. Time use and food intake have limited variation and can use fairly small samples. Total consumption and total food consumption also have limited variation, while other consumption and detailed consumption have bigger variation. The needed precision also differs. Consumption and labour force data have high priority for the survey and the sample should be big enough to give robust estimates for those. Generally 200 – 300 households are adequate for presenting estimates on a group of households. Any affordable sample will be too small to allow for estimates on district level. The aim for the INCAF is to provide estimates on provincial level, with rough estimates for a few break downs for each province. Common sizes for consumption surveys in Europe are 3 000 – 10 000 households (with only 1 570 in the Netherlands and more than 50 000 in Germany) and labour force surveys 10 000 – 50 000 people (ranging from 3 100 in Iceland to 134 000 in Italy) around 2005. The European Survey on Income and Living Conditions (EU-SILC) is commonly 4 000 - 9 000 households (ranging from 3 800 in Cyprus and Latvia to 16 000 in Poland). Portugal had 10 400 households in HBS, 36 500 people in LFS and 5 000 households in EU-SILC.

A proposed sample of 8 652 households and about 34 000 people for LFS is well above the European average. The INCAF team will try to spread the sample on more EAs in bigger cities. This will depend on the team's ability to update/list more EAs.

Província	TOTAL		URBANO		RURAL		Household Sample Urbano	Sample Rural AF	Amostra Global AF
	Agregados Familiares	UPAs	População	UPAs	População	UPAs			
Niassa	252 550	64	267 511	30	901 793	34	390	340	730
Cabo Delgado	391 271	66	332 689	32	1 270 951	34	416	340	756
Nampula	920 821	77	1 137 290	40	2 844 858	37	520	370	890
Zambézia	874 589	71	670 868	34	3 176 362	37	442	370	812
Tete	379 974	64	242 928	30	1 539 428	34	390	340	730
Manica	269 717	64	356 284	30	1 054 217	34	390	340	730
Sofala	321 051	78	626 925	44	1 013 804	34	572	340	912
Inhambane	273 555	64	281 592	30	989 192	34	390	340	730
Gaza	240 535	64	312 074	30	915 453	34	390	340	730
Maputo - Província	248 030	78	817 278	44	385 340	34	572	340	912
Maputo - Cidade	210 798	120	1 092 472	120	0	0	720	0	720
Moçambique	4 382 891	810	6 137 911	464	14 091 398	346	5 252	3 460	8 652

The design has not only to take care of variation between households, but also variation over time and type of consumption. There is variation within a month with much higher consumption after salary payments than the weeks before and there is variation between seasons, particularly for the rural population. Visiting a household the first week in a month in the first quarter, the second week in a month the second quarter, etc. will take care of those variations and give a more representative picture for each household for more accurate poverty classification.

With four visits to each household, monitoring 1 week non-durable goods, 3 months semi-durable goods at each visit and 12 months durable goods at first and 9 months at the last visit gives following total number of consumption months for a proposed INCAF sample of 8 652 households, compared to the latest HBS with 10 800 households. The INCAF design is much more efficient compared to the traditional HBS. Semi-durable and durable goods have much bigger variation and are the weakest when it comes to reliability. Consumption months are a proxy for data quality.

Measured months as proxy for reliability

	INCAF	HBS 2008/2009
Non-durables	8 652/17 304*	2 700
Semi-durables	103 825	32 400
Durables	103 825/181 692*	126 000

* The second year with HBS module included

The sample in INCAF will also be spread on more enumerator areas (EA), which is more efficient from a quality point of view and will compensate for the some smaller sample size.

The panel approach will improve the quality of trend data, e.g. quarterly change in employment rate, but also give better information on long-term unemployment as the same people are followed for one year.

As poverty development is a priority for the INCAF, the inclusion of weekly as well as seasonal variations for the same household will limit the randomness in poverty classification, which must be quite big in previous HBS, where food consumption was based on only one random week per household. If you visit a well off household the week before salary it may well be classified as poor, while visiting a better-off farmer before the harvest it may be classified as poor.

It is proposed that the same PSUs are used at least the second year. It will save time and have a partial panel effect. The saved time can be used for the second visit per quarter to collect diaries, if an HBS module will be added.

The second sampling unit is dwellings rather than households, allowing newcomers in a dwelling as a substitute for households who have moved out. Substitution for non-respondents should not be practiced, as it will improve the sample error only marginally, but risks introducing bias.

Rotation of the panels is optional. Common in LFS is to let each household to be included five quarters and therefore to change 20 % each time. This improves change data not only between quarters but also from year to year. Also practiced and maybe simpler is to change 25 % each quarter. Some countries change the whole sample in the beginning of the year also in LFS. If the new households are sampled in the old PSUs it will at least be a semi-panel effect between years. A rotating scheme can be more complicated when the HBS module is added giving a longer period for the module, arguing some for introduction of rotation after the first two years. INE will continue to discuss the options and will decide later.

The plan was to have to teams in each province, one for urban and one for rural areas. It is probably more efficient if both teams can work in both areas, allowing more flexibility.

IT solution

There are mainly two alternatives for the data management:

- 1 CPro (or Blaise) + SPSS, STATA or SAS
- 2 MS Access eventually combined with SQL.

INE has more experience in the first alternative, but it has some weaknesses.

Database design

As INCAF will be a panel survey with visits four quarters and with many different areas to be monitored, the database will be rather complex with more than 40 separate data sets. Most sets should be possible to link. A relational database (like MS Access or SQL) would be easier to manage. It is then easy to build and maintain the links without any data manipulation.

Q 1	Q 2	Q 3	Q 4
EA data	Code lists		
ID data	ID data	ID data	ID data

Household comp	Housing	Child labour	
Person information	Health	Structural labour force	Quality of life
Employment	Employment	Employment	Employment
Non-durable cons.	Non-durable cons.	Non-durable cons.	Non-durable cons.
Own-produced food	Own-produced food	Own-produced food	Own-produced food
Semi-durables	Semi-durables	Semi-durables	Semi-durables
			Durables
Consumer confidence	Consumer confidence	Consumer confidence	Consumer confidence
Tourism	Tourism	Tourism	Tourism

Different from alternative 1, the tables with controls and look-up tables (code lists) are first designed in the Access database, then the forms can either be done manually and linked to the tables where data will be stored or Access can automatically create the forms with the links for further design if wanted. Data are never stored in the forms and therefore never moved. The real strength with Access is that it is strong in database management and data can be managed from entry to tabulation without being moved or separated manually.

Data entry

The data will be entered on tablets in the field and most of the data controls should be computerized. There are mainly three alternative programs to use: CPro, MS Access and Blaise. All programs have capacity to control data. Blaise is built for Computer Aided Personal Interviews (CAPI), but is expensive and not known by INE. CPro has been used e.g. in the Demographic and Health Survey and in the INCAF pilot. It is free of charge and is built to enter data for censuses and surveys. It is, however, not designed for CAPI and mainly for simpler surveys with few data sets. MS Access is almost free (as part of MS Office Professional). It is robust and, though not directly designed for surveys, one of the most used data entry programs (as engine) for accounting systems in small and medium sized companies. It is also designed for easy use on internet, with ready-made objects for the screen and skips capabilities.

Data communication and storing

Data are planned to be sent in by email. Access has both readymade functions for sending and achieving data by email and storing automatically in a central Access database or an SQL database.

It is important to always make back-ups in three steps

- 1 Raw-data from the field before any editing, coding or other manipulations
- 2 Manipulated data
- 3 User file/data matrixes

Editing and coding

Most of the data editing will be done with computerized controls at the data entry. Most coding will also be done at the interviews with ready-made code lists. Some variables, like e.g. occupation are too complex to manage in the field and have to be done at central INE. Some more editing will also be done, particularly checking the consistency between the data sets. It is advisable to limit the controlling to be efficient and not delaying the results too much. Earlier it was common to make all possible consistency and value checks, but studies in e.g. Canada and Sweden have showed that they often have limited impact on the results. It is important to know that different data need different quality. Identification data should be perfect to guaranty the links between data sets. Data which are used for classifications are also important, as mistakes affect both the categories where they will be and the categories they rightly belong to. It means that the data about the head of the household is more important than for the others in the household. It is also important to check out-layers (extreme data) to find if they are true or not. Much of

the error will be avoided by the data entry in the field, as all the mentioned programs have good checking capabilities.

Data manipulation

Data from CSPro have to be separated to the different tables by using SPSS or STATA before the manipulations. This needs both computing skills as well as good knowledge about the data. It takes time and is risky. In Access/SQL, data are already stored in the respective tables for further manipulations (extracting, aggregating, weighting and normalizing periods) as well as creation of new variables. In Access the manipulated data can be stored in new tables or just being queries keeping the original data and only manipulating them when the tabulation is made. Any change in the original data will then be followed through to the final tables.

Tabulation

All standard tables can easily be created in SPSS, STATA as well as with the pivot table function in Access. Most household and individual data will be tabulated as frequencies while the economic data will be presented as total consumption for the country but broken down by all consumption items and as averages for all households and groups of households.

Files for internal use and researchers can be stored as tables with each household (or person) as a row and the variables (and household consumption for the 12 COICOP groups as columns). By aggregating households with same classification variables, anonymous users' files can be available on internet for further tabulations in e.g. Excel.

Questionnaire design

Household characteristics

Household characteristics will only be asked at the first visit. Alternatively, changes can be included in following visits by asking for new members or members permanently moved in or out or died. For socio-demographic statistics the first alternative is adequate. New born babies' impact on the economic data should be marginal. The number of adults has at least an impact on household economy, only marginal for macro-economic statistics, but possibly for poverty analysis. Movements in and out between households as well as dead have probably only marginal impact on household averages similar to the impact of visitors, which most countries don't include anyway.

Housing

The housing questions could be reduced later to focus only on factors correlated with imputed rent. One such factor is major damages, which should be added now.

Health services

The health service module is proposed to come later as a separate extended module about health and health services not covered by DHS. Service performance and other qualitative aspects can also be measured in an eventual SPQL module. Health expenditure will be measured quarterly as semi-durables.

Tourism module

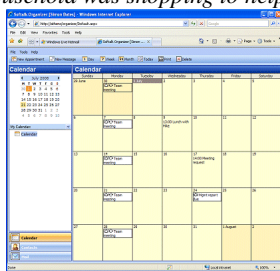
Tourism is more a household activity than an individual. It is recommended to be measured trip by trip instead of by individuals. A formulated question for the expenditures is needed. If the module is not limited to domestic trips it will be possible to get some estimates for cross border shopping.

Non-durables

It is strongly recommended to start the non-durables with a lump sum question about for how much of non-durables were bought in total the past seven days. This is easier for the household to remember than what they spent in details for different things. The list of detailed consumption should be shorter and asked as a part of the total sum. It is very important to tell the respondent that the “today will not be included” and “we start with last ...day”. Otherwise it is a high risk that many answer for eight days.

I will now ask about your expenditures. Include all household members but exclude what is bought for agriculture or business.

I will start to ask about the day-to-day consumption like food, hygiene products and cleaning materials. I want you to include all that the household bought during the last seven days. Don't include today. Seven days ago started onday. *Cross the days that should be included and ask for the days the household was shopping to help to memorize.*



1 Did anyone in the household during the past seven days buy...	For how much in total Mt)?
...food, drinks, hygienic products or other consumables (paper, matches, cleaning and washing materials, etc., but not gasoline, fuels	
How much of that was for xxxxxxxxxxxxxxx ²	
And how much was for yyyyyyyyyyyyyyy	
Meals and drinks outside home	

Own produced food

This part took a long time in the pilot. It was agreed to only measure yesterday, as most rural households eat similar staple food most days, though it can be different between areas and seasons. Fish and meat should be added to the list.

Semi-durables

Semi-durables, like clothing, should be asked for the past **three** months (at each visit), as it was done in the previous HBS. The list will be adjusted some and include health service costs. The purchase of second hand articles are not marginal in Mozambique and should be asked for, by asking if it was bought new or second hand. If it is common that the household are buying both new and second hand of the same kind, e.g. clothing, it is better to ask “of this value, how much was second hand?”

All consumption should be included, so each group should have “other...”, e.g. other cost for vehicles.

If cross border shopping not measured by the custom is common, it could be asked how much was bought abroad. An alternative is to measure this in the tourism module, if the trips include foreign trips.

² If needed by e.g. NA a few or some specifications can be asked, but they must be specific, not aggregates

Durables

The possessed and bought numbers are not important but new and second hand should be included also here. Tablets (like iPad) should be added to computers.

Labour force and Child labour

The labour force and child labour modules should be integrated, but separated into quarterly, annual (structural), and rotating indicators.

HBS module

The second year an HBS module is proposed, using supplementary diaries and extended list of semi-durables for detailed consumption. Questions about height, weight and physical activities should be asked to allow estimation of BMI and establish more detailed calorie needs.

Community module

Local prices will be collected to support the estimation of the value of own produced food. Other community relevant indicators (e.g. access to services) could be added.

Other modules

Example questionnaires for food intake, time use, informal sector, health, children's well-being and migration and remittances have been handed over in electronic form.

Welfare/poverty monitoring

Poverty line

The light INCAF will not have necessary information for creating a new poverty line, so the latest should be updated with CPI. A new poverty line could be decided after the second INCAF year, when detailed information is available about the food consumption. Then also data (height and weight) could be collected to calculate individual Body Mass Index (BMI). If individual activity level is asked it would be possible to estimate individual calorie needs. The daily per capita need for Mozambique is set to 1,800 kCal by WHO and FAO taking into account the age structure and based on BMI. New scales of equivalence have to be developed in close cooperation with the responsible for PARP.

Poverty headcount and poverty rates

Though the data will be much better suited for money-metric poverty assessment, compared to previous HBS, it will probably not be possible to measure changes in poverty rate from one year to the other with statistical significance. Other short term welfare/poverty indicators should be used. One simple, but robust indicator is the food share of total consumption.

Engel's Law

"The poorer the family, the greater the proportion of its total expenditure that must be devoted to the provision of food. . . .The proportion of the expenditures used for food, other things being equal, is the best measure of the material standard of living. . . ."

{Ernst Engel (1861)}

This simple measure is very robust and does not need local prices, kCal or scales of equivalence. Following table from the HBS 2008/09 show how robust it is as group estimate. Though better off households put more money on food, the food share declines.

Household group	Total consumption Meticais/month	Food consumption Meticais/month	Food share %
Urban	5,333	1,876	35
Rural	2,466	1,621	66

Norte	2,876	1,629	57
Centro	2,723	1,742	64
Sul	4,953	1,710	35
Maputo Cidade	10,497	2,262	22
Nenhum	2,251	1,524	68
Primario do 1 grau	2,725	1,641	60
Primario do 2 grau	3,808	1,880	49
Secundario e mais	7,151	2,087	29
Outros	7,279	1,952	27
Quartil 1	771	484	63
Quartil 2	1,491	997	67
Quartil 3	2,193	1,497	68
Quartil 4	3,262	2,119	65
Quartil 5	8,780	3,376	39

Only the quintiles are not following the pattern. Even quintile 5 should have lower rate (even lower than Maputo Cidade). The reason is probably because the classification is done by consumption per household. Such classification must be done by adult equivalents. Now big, but not wealthy households are mixed with small but better off households at each level.

The food share can even be used as a quarterly indicator/index for welfare/poverty.

Dissemination

It is very important to come out with results as soon as possible, so any possible bottleneck in the data flow has to be solved. Otherwise it will have a stopping impact on coming data. But more important is to demonstrate the capability of the survey to please the users with data when needed.

The old culture of releasing by reports should be changed. Basic quarterly and annual results should be fairly automatic and released within three months on internet. More analytical reports can be produced later by separate subject matter staff or outside institutions.

Risks and solutions

Most countries are able to collect data in a fairly efficient and accurate way. The bottlenecks are most common in the central operations of the statistical process. They are also most sensitive to mistakes. Mistakes in the field are often random or limited. Data entry can however easily be biased, but can be limited by built-in controls in the tablets. Mistakes in the weighting procedure can also have a big impact on the results. The advantage with a continuous survey is that the same procedures can be used and improved.

To be successful and having a positive impact on INE's image and confidence for statistics it is important that INCAF is prioritized and have its own resources all the way from planning and design to release on internet.

Another main risk is "special" and often donor driven surveys not willing to integrate with the INCAF, robbing (and often overpaying) available limited resources.

In the beginning it will be an adjustment period to facilitate different needs, but it is important to rather soon come to a national (but following international recommendations) set of standard background variables and core indicators (e.g. whether measuring access to service in time or distance). This is to facilitate comparability over time and space (national, regional and globally).

A third risk is overloading. INCAF can easily facilitate many different needs, but not everything at the same time. The core questionnaire should be kept fairly clean and not being contaminated by odd or "good to know" questions. Every year the need for continuing a question or bringing in a new should be carefully scrutinized using following principles:

- Have the results been used (who is using e.g. marital status)?
- Is there a defined use for a new question?
- Is the question possible to answer?
- Has it successfully been used somewhere else or is recommended internationally)?

- Will it generate significant results (poverty development is hardly significant from one year to another)? Is the group of interest big enough compared to the sample size?

Same data should not be collected by parallel surveys. It is inefficient use of resources and will probably create contradicting results, bringing down the confidence for all statistics. At worst, some households will have to answer the same question twice and judge INE as a bureaucracy that does not know what it is doing.

If the INCAF cannot continue of any reason, the costs are well invested. If, e.g. the HBS is decided to be separate, the first INCAF year can be seen as the first full scale LFS according to ILO requirements with continuous measurement of the labour market, plus a lot of other information. Investments in training and equipment (about 1 MUSD) can be reutilized for the HBS.

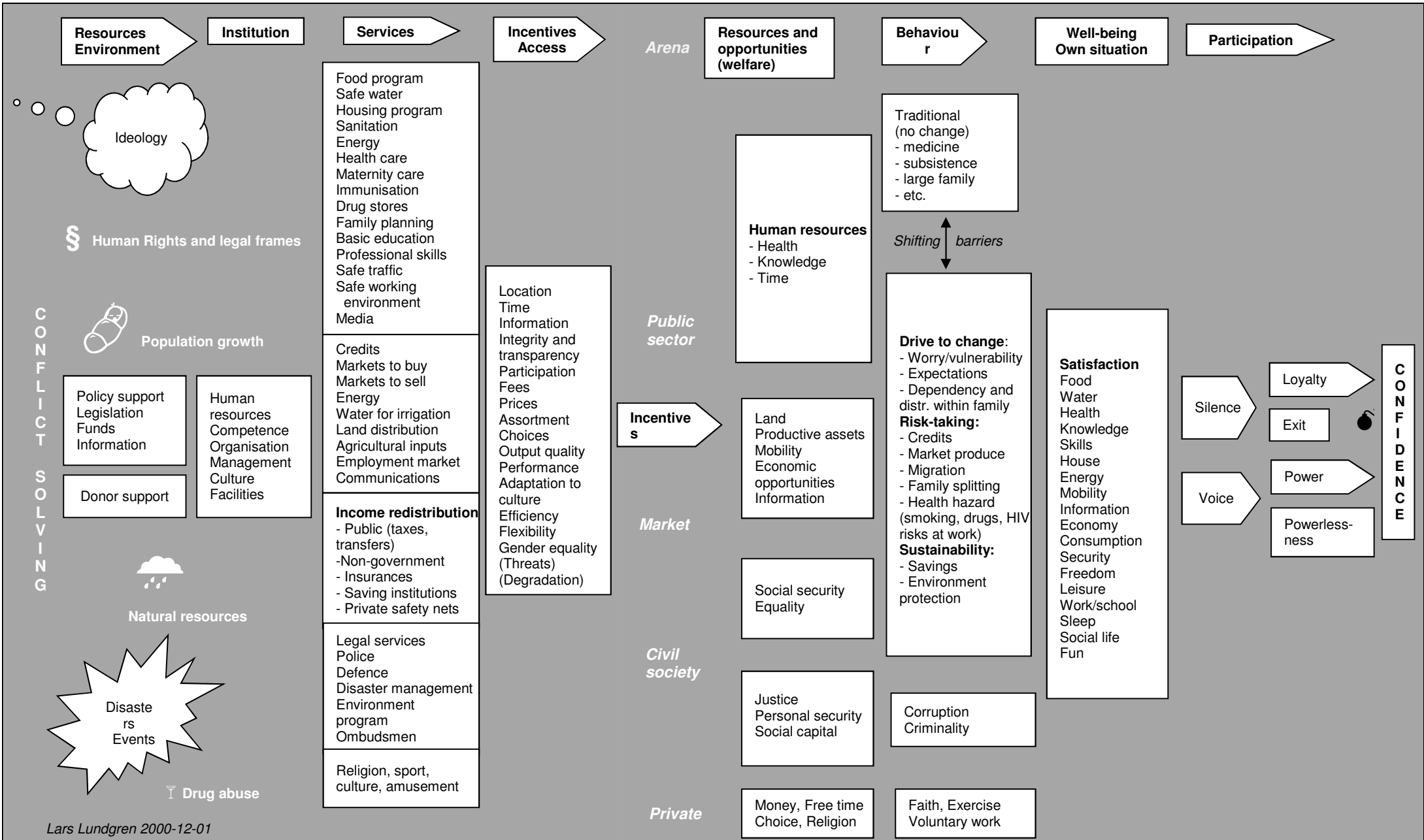
Documentation

Documentation and metadata are very important for any survey. Attached in annex 3 is a template for a minimal documentation of INCAF, based on EUROSTAT's requirements for HBS.

The manuals for interviewers should, as much as possible, be integrated in the questionnaire/tablet. With most important instructions on the screen and others accessible by a help-key



Annex 1: A mean-end (cause-effect) map for well-being



Annex 1: A mean-end (cause-effect) map for well-being

Annex 2: Example of SPQL questionnaire

Household No. <input style="width: 100%;" type="text"/>	Person No. <input style="width: 100%;" type="text"/>	First name <input style="width: 100%;" type="text"/>
age 18-64		

Service Performance Quality of Life Index

The questions refer to your own personal conditions during the last year. Please rate the different conditions/ circumstances on a scale from one (1) to ten (10). One means *not at all satisfied* and ten means *extremely satisfied*. If you have no experience of a question, circle *No opinion* (?). **Answer more by feeling** and without further explanation.

	<i>Not satisfied</i>		<i>Very satisfied</i>		No opinion
1 Personal conditions	☹	☺	☺		
How do you rate your...					
- access to safe food	1 2 3 4 5 6 7 8 9 10				?
- access to safe drinking water	1 2 3 4 5 6 7 8 9 10				?
- health on the whole	1 2 3 4 5 6 7 8 9 10				?
- capability to read and write	1 2 3 4 5 6 7 8 9 10				?
- professional skills	1 2 3 4 5 6 7 8 9 10				?
- safety from burglary	1 2 3 4 5 6 7 8 9 10				?
- safety from violence	1 2 3 4 5 6 7 8 9 10				?
- safety from corruption/extortion	1 2 3 4 5 6 7 8 9 10				?
- influence at home	1 2 3 4 5 6 7 8 9 10				?
- influence outside home	1 2 3 4 5 6 7 8 9 10				?
- trust in politicians	1 2 3 4 5 6 7 8 9 10				?
- trust in professionals in the public sector	1 2 3 4 5 6 7 8 9 10				?
- trust in the business sector	1 2 3 4 5 6 7 8 9 10				?
- trust in the community you live in	1 2 3 4 5 6 7 8 9 10				?
- trust in the judicial system	1 2 3 4 5 6 7 8 9 10				?
How do you rate...	☹	☺	☺		
- your life now as a whole	1 2 3 4 5 6 7 8 9 10				?
- your living conditions as a whole	1 2 3 4 5 6 7 8 9 10				?
- your life compared to other people	1 2 3 4 5 6 7 8 9 10				?
2 Housing and environment - How do you rate the...	☹	☺	☺		
- sanitary standard at home	1 2 3 4 5 6 7 8 9 10				?
- access to water at home	1 2 3 4 5 6 7 8 9 10				?
- access to energy for lightning and cooking	1 2 3 4 5 6 7 8 9 10				?
- government support for better housing (land, credits, etc.)	1 2 3 4 5 6 7 8 9 10				?
- access to internet	1 2 3 4 5 6 7 8 9 10				?
3 Economic opportunities - How do you rate...	☹	☺	☺		
- your possibilities to have/get a decent job	1 2 3 4 5 6 7 8 9 10				?
- your possibilities to get commercial credits	1 2 3 4 5 6 7 8 9 10				?
- your income from work	1 2 3 4 5 6 7 8 9 10				?
- the general price level	1 2 3 4 5 6 7 8 9 10				?
- the legislation and regulations for doing business/farming	1 2 3 4 5 6 7 8 9 10				?
4 Social security - How do you rate the...	☹	☺	☺		
- public safety net (pensions, social transfers, free services)	1 2 3 4 5 6 7 8 9 10				?
- support from other sources (relatives, employers, organisations, etc.)	1 2 3 4 5 6 7 8 9 10				?

Annex 2: Example of SPQL questionnaire

5 Services - How do you rate your access to high standard...	Not satisfied		Very satisfied		No opinion						
	☹	☺	☺	☹							
- health care	1	2	3	4	5	6	7	8	9	10	?
- ART service	1	2	3	4	5	6	7	8	9	10	?
- medicines	1	2	3	4	5	6	7	8	9	10	?
- birth attendance	1	2	3	4	5	6	7	8	9	10	?
- vaccination programs	1	2	3	4	5	6	7	8	9	10	?
- training to develop your professional skills	1	2	3	4	5	6	7	8	9	10	?
- education for your children	1	2	3	4	5	6	7	8	9	10	?
- leisure facilities and activities	1	2	3	4	5	6	7	8	9	10	?
- local transportation	1	2	3	4	5	6	7	8	9	10	?
- regional transportation	1	2	3	4	5	6	7	8	9	10	?

Following indicators try to measure your own willingness to change, if the public sector would support you.

6 Changes - How interested are you for...	Not interested		Very interested		No opinion						
	☹	☺	☺	☹							
- earning money	1	2	3	4	5	6	7	8	9	10	?
- give your children high education	1	2	3	4	5	6	7	8	9	10	?
- immunise your children	1	2	3	4	5	6	7	8	9	10	?
- save water	1	2	3	4	5	6	7	8	9	10	?
- protect the environment from degradation	1	2	3	4	5	6	7	8	9	10	?

7 Working conditions – only for employees	Not satisfied		Very satisfied		No opinion						
	☹	☺	☺	☹							
How do you rate your											
- wage?	1	2	3	4	5	6	7	8	9	10	?
- working environment?	1	2	3	4	5	6	7	8	9	10	?
- influence in work decisions?	1	2	3	4	5	6	7	8	9	10	?

8 Conditions for agriculture – only for farmers	Not satisfied		Very satisfied		No opinion						
	☹	☺	☺	☹							
How do you rate the											
- output/harvest from the agriculture past 12 months?	1	2	3	4	5	6	7	8	9	10	?
- output/harvest from the agriculture previous 3 years?	1	2	3	4	5	6	7	8	9	10	?
- expected next output/harvest?	1	2	3	4	5	6	7	8	9	10	?
- protection of crops?	1	2	3	4	5	6	7	8	9	10	?
- compensation for lost harvests?	1	2	3	4	5	6	7	8	9	10	?
- access to land?	1	2	3	4	5	6	7	8	9	10	?
- access to fertilizers, pesticide, etc.?	1	2	3	4	5	6	7	8	9	10	?
- access to markets or distributor for selling products?	1	2	3	4	5	6	7	8	9	10	?
- prices offered for the products?	1	2	3	4	5	6	7	8	9	10	?
- government services for agriculture (xxx,, etc.)?	1	2	3	4	5	6	7	8	9	10	?
- fees for such services?	1	2	3	4	5	6	7	8	9	10	?

9 Conditions for business – only for household business	Not satisfied		Very satisfied		No opinion						
	☹	☺	☺	☹							
How do you rate the											
- revenues from the business past 12 months?	1	2	3	4	5	6	7	8	9	10	?
- revenues the previous 3 years?	1	2	3	4	5	6	7	8	9	10	?
- expected revenues coming 12 months?	1	2	3	4	5	6	7	8	9	10	?
- access to credits for business?	1	2	3	4	5	6	7	8	9	10	?
- access to markets or distributors to sell products?	1	2	3	4	5	6	7	8	9	10	?

Annex 2: Example of SPQL questionnaire

Annex 3: “minimum” quality report

Table 1 Overall information

Title of the survey	INCAF
Reference years	2012/2013
Frequency	Continuous
Achieved sample size	Total number of households 8652
EUR/NAC conversion factor	
NAC/PPS conversion factor	

Table 2 Main sampling characteristics⁶

Ultimate sampling unit	The survey collects information on households and persons as well. The ultimate sampling unit (USU) is “built” around the selected dwellings.
Probability sampling	Yes
Number of sampling stages	Two-stage selection: First, a systematic sample of 810 geographical enumeration areas is selected, with probabilities proportional to size. The second stage consists of a simple random selection of 10 households within each sample urban area, 6 in cities and 13 in other urban areas.
Stratification criteria	Urban/rural and provinces
Over-sampling of special domains	Urban areas
Survey population: main exclusions	Collective or institutional households (old persons’ homes, hospitals, hostels, boarding houses, prisons...) are excluded, as in practice homeless people.
Sampling frame	Population census 2007. PSU updated before sampling households.
Whether substitutions are allowed	Not allowed

Table 3 Sample size and non-response errors

Gross sample size	8 652
Number of eligible units	
Number of units (USU) successfully contacted – before and after substitution	
Number of responding households – before and after substitution	

Annex 3: “minimum” quality report

Table 4: Weighting

Calculation of the household design weights	Calculated for each sampling unit as the reciprocal of the probability of selection of the unit
Weight adjustments for non-response at household level	The weights are adjusted for non-response at household level by $10/n$ ($6/n$ or $13/n$), where n is the number of responding households in the EA
Weight adjustments to external data sources (calibration)	Not yet. In most cases, the calibration technique makes the accuracy better, therefore, is widely used to improve the quality of survey data.
Any other weight adjustments	Adjustments for missing visits for responding households and to have same representation of all weeks.

Table 5: Estimated standard errors, confidence intervals and design effects

Indicator	Mean total household consumption expenditure and one-digit COICOP categories; age of the household’s reference person: less than 30, 30-44, 45-59 and 60+ years; household type: single person, two adults, three adults or more, single parent with dependent children, two adults with dependent children, three or more adults with dependent children; Employment status of the household’s reference person: manual worker in Industry and Services, non-manual worker in Industry and Services, self-employed, unemployed and other inactive
Achieved sample size	
Estimated value	
Estimated coefficient of variation (%)	
95% Confidence interval – lower bound	
95% Confidence interval – upper bound	
Estimated design effect (Deff)	

Table 6: Household, household membership and head of household

Household defined as persons sharing	Accommodation	Y
	Expenditure	Y
	Income	Y
	Family emotional ties	N
	Other (indicate)	
Household membership	Usually resident, related to other members	Y
	Usually resident, not related to other members	Y
	Resident boarder, tenant, lodger	N
	Visitor	N
	Live-in domestic servant, au pair	Y
	Resident, absent from dwelling in the short-term	Y
	Children in household, in education away from home	Y
	Long-term (>12 months) absence with household ties: working away from home	N
Temporary absence with household ties: in hospital, nursing home or other institution	Y	
Head of household (indicate)	The person designated as such by the household concerned. Other persons (with highest income, highest education etc.) can be used as reference person	

Annex 3: “minimum” quality report

Table 7: Consumption expenditure approach

Consumption expenditure approach	Actual final consumption	Y
	Final consumption expenditure	Y
	Monetary consumption expenditure	Y
	Other (indicate)	
	Consumption expenditure approach: Main exclusions	
	Reference periods for expenditure	4 weeks for non-durable and 12 months for semi-durable goods
Borderline cases: recording and valuation	Goods or services for own final consumption	Valued at market price
	Leasing and hire purchases	Valued at full price at acquisition
	Health and education expenditures	Actual (net=gross) expenditure (no reimbursements)
Estimation of imputed rentals: population	Owner-occupiers: Principal dwellings	N (but 2014)
	Owner-occupiers: Secondary dwellings	N
	Tenants: reduced or provided rent-free	N (but 2014)
Estimation of imputed rentals: method	Self-assessment	Y
	Stratification	(Y)
	Log-linear regression	N
	Heckman regression	N
	Other (indicate)	
	List of the variables used in the model:	
Salaries and wages in-kind: which benefits are evaluated	Free or reduced cost housing	(2014)
	Private use of company car	N
	Gas, electricity or water	N
	Telephone	N
	Other (indicate)	

Annex 3: “minimum” quality report

Table 8: Data collection

Diaries	Recording unit	Household
	Recording period	One week in 4 quarters
	Distribution of recording periods over the survey year	1/12 of the annual sample is surveyed each month
	Items covered in the diary	Non-durable and semi-durable goods
Collection of substantive information	Instrument	
	Recording unit	
	Items covered	