



Inception Report

January 26 2005

TA for the Scandinavian Support Program to Strengthen the Institutional Capacity of the National Statistics, Mozambique

Karsten Bormann



Instituto Nacional de Estatística

*Karsten Bormann
Instituto Nacional de Estatística
Av. Ahmed Sekou Touré 21, 5
Caixa Postal 493
Maputo
Mozambique
karsten.bormann@ine.gov.mz
++258 84 56 86 88*

Table of contents

1	Executive Summary	5
2	Introduction	6
3	Urgent Activities	7
4	Strategic Duties.....	8
4.1	IT management.....	8
4.2	Quality Assurance	9
4.3	Data warehouse	9
4.4	Provincial offices.....	10
5	Appendix 1. Terms of Reference.....	11
6	Appendix 2. Scouting for resources at the home office	13
6.1	Summary of findings	13
6.2	Mission background	13
6.3	Findings.....	13
6.3.1	Data Warehousing at INE	
6.3.2	The INE Portal	
6.3.3	The migration of the INE server environment	
6.3.4	Potential missions	
6.4	Conclusion and Outlook.....	17
6.5	Mission TOR	17

List of abbreviations

CO	Scanstat Coordination Office in Statistics Denmark
Danida	Danish International Development Assistance
DKK	Danish Kroner
DSt	Statistics Denmark
EUR	European Euro
INE	Instituto Nacional de Estatística, Mozambique
INE-P	Instituto Nacional de Estatística, Portugal
MZM	Mozambique Meticaís
NOK	Norwegian Kroner
Scanstat	Consortium between Statistics Denmark, Statistics Norway and Statistics Sweden
SCB	Statistics Sweden
SEK	Swedish Kronor
SSB	Statistics Norway
USD	US Dollars
ZAR	South African Rand

1 Executive Summary

The following document attempts to outline the planned duties of the long term advisor within the area of IT and statistical production. The work has two components namely helping suggest and design technical solutions and improving management of our human and technical resources in order to consolidate the gains already experienced in the IT area.

2 Introduction

The present long term advisor replaces Mogens Grosen Nielsen as advisor in the area of IT and statistical production. The primary local counterpart for this position is the head of the IT department (DISI), Anastácia Honwana. Work proceeds in close collaboration with her as well as with the technicians involved in a given task.

The activities naturally follow up on the final observations and recommendations of Mogens Grosen Nielsen, allowing for the fact that conditions and priorities change and adding activities in areas where the incoming advisor feels that it is advantageous.

The aim of the Scandinavian program and, by inference, having a long term consultant within the IT area is that of capacity building. Ideally, most of the effort then goes into strengthening the IT organisation with the aim of being able to provide a constant and reliable service. From such a consolidated basis one may then expand activities towards strategic goals to the extent that the resources suffice.

3 Urgent Activities

Large part of the initial three months were dedicated to trying to bring under control areas where we had problems delivering basic services. Three areas needed urgent attention, namely; the INE portal, the DPINE mail, and the stability of the server environment.

- DPINE email: The task here has been to make email available at (all) the DPINEs. We are currently moving towards a solution that is less sensitive to the quality of the connection between the DPINEs and INE. The implementation is to start at the DPINEs during January 2005.
- INE portal: The portal's response times needed to be lowered (across all ISPs). We have been trying to minimise the webpage sizes and have tried to consolidate the current solution (initially with the aim of hosting the server outside Mozambique), as well as doubling the bandwidth. Response times are now deemed acceptable.

Server environment: In order to minimise software conflicts and to be able to focus competency building within fewer areas, which we then should be able to master better, the servers will be migrated to run on the same operating system (Windows 2003 Server) in March 2005. The migration process have been discussed and partially planned with Bo Guldager from Statistics Denmark, and steps have been taken to prepare the staff through course participation.

4 Strategic Duties

4.1 IT management

INE has been following an ambitious path within the IT area, which in turn has stretched resources thin. Thus there is a need to consolidate the organisation around the current services to ensure a continuous and stable service. The need becomes even bigger as the most ambitious of the strategic priorities is still ahead of us; building a common architecture for all of INE's data processing around a data warehouse.

In order to provide a more informed basis for helping us consolidate the IT operations, and to ascertain that we have adequate resources both as regards the number of hands and their skills, to execute the strategy, we conducted a survey that asked the DISI staffers about their current tasks and how well we are currently executing on these tasks. More specifically;

- Current tasks and the resulting products/outcomes
- How vulnerable is the operation/product?
- What are the consequences of failure?
- Who makes decisions about introducing tasks and choosing paths of action?
- Any outstanding problems or major issues?
- Anything that could/should be improved? How easily can it be done?
- Who assigns you to tasks?

Furthermore, we asked DISI members to detail their

- Competencies, skills, experience, and (long term) interests.

Besides providing an overview of IT operations it is the intention that we, during the first half of 2005, use the survey results to undertake the following tasks and improvements:

- SWOT analysis of DISI: We are of course very interested in the threats to the current services but, for a more strategic approach, it is important to get tasks prioritised in order to free up resources that can then be used for consolidating the most important current services and for the implementation of strategic goals.
- Implementation of the 3x3 rule: The 3x3 rule states that all important tasks should be covered by at least three staff members (to allow for a consistent service level) and that all staff members should cover at least three important task areas (for the organisation to be flexible).
- Creation of 'self-sustainable' teams within important task areas: With major tasks covered by at least three people that can corroborate the skills of one another the teams should be able to function independently and without much reliance on outside help. It is envisaged that the teams take responsibility for their own continued skills enhancement by studying books and magazines and doing hands-on experimentation, as well as sharing the knowledge gained. From the management side this effort must be supported by allocating funding for books and magazines, and test equipment. And, harder; allocating the time needed for the staff members to engage in competency building.

- Knowledge sharing between subject teams is envisaged to take place by the groups making small seminars during DISI's Monday morning meetings.

4.2 Quality Assurance

A major part of the recommendations of the outgoing long term consultant regards quality assurance procedures which have been thoroughly discussed from a birds-eye viewpoint but which all but lack implementation. It is the plan to introduce QA procedures to new developments during 2005. Some focus areas identified are:

- Implementation of quality assurance procedures: A schematic view of the process steps involved in application development at INE was created by the previous consultant and DISI's quality group. Each of the steps are to be corroborated with templates for the documentation to be created. A first step was taken in the creation of a test specification for the DPINE mail solution (and the corresponding test).
- Get 'test systems' for the staff to experiment with: To comply to quality assurance procedures it must be possible to try out solutions before applying them to a production environment. Currently, test machines for this purpose are lacking. Reshuffling of servers during second quarter of 2005 may remedy this situation somewhat. Subject to the approval of the suggested budget for the IT subproject the situation may be alleviated still further.
- Disaster recovery procedures: Due to budgetary constraints physical duplication of systems is not feasible and we will settle for a proper description of server configurations. A first step was taken in the process of preparing the portal for being hosted outside INE. (This process was later aborted but the exercise was a sound one). Next, the configuration steps should be described in detail when the servers are migrated to windows 2003. Also, steps to store data and recovery procedures outside INE should be taken
- Follow-up on recommendations of short term missions probably should be made a little more formal.

4.3 Data warehouse

To streamline the statistical production flow the INE strategy calls for the introduction of a data warehouse as the central pillar of the (statistics production part of the) IT environment. This is a major effort, and even if it is attempted jumpstarted in February 2005 it must be expected to take all of 2005 before a prototype is up and running and all of 2006 before all production processes have been rearranged around the data warehouse. Success will hinge on the ability to free enough resources for the effort as well as the enthusiastic involvement of the statistics departments whose data must be represented in a form adequate for their purposes - and without collisions of nomenclature between departments. The tasks identified so far are:

- Write a data warehouse awareness paper to show the role of the data warehouse and the reasons for the expected gains in production efficiency. This task was completed in the fall 2004.
- Architectural design of the data warehouse and the infrastructure. Work was done in the fall 2004 with Clara Panguana and discussions continued with Soeren Netterstrom in Copenhagen. It is the plan to more or less finalise the overall architecture in cooperation with DISI's applications group during the data modelling workshop:
- Data modelling for data warehouses, workshop with Soeren Netterstrom, Statistics Denmark (week 6, 2005): The critical part of the effort is to get the metadata right. We need to ensure that we have a common language and reference frame for the discussion of problems and that we have some understanding of the modelling techniques normally used for data warehousing. This workshop attempts to provide this and to jumpstart the data warehousing effort.
- Make a compilation of deliverables, data sources and the resulting data, and the software tools that must be supported by the data warehouse implementation.
- Scouting for technologies that allow a low cost implementation. (Work in progress).
- Data warehouse implementation.

4.4 Provincial offices

The IT services at the provincial offices (DPINEs) have not been sufficiently consistent. The email situation has already been mentioned. But heading for the root causes of the problems we should attempt to:

- Ensure that the IT-infrastructure of the provincial offices be made simple and homogeneous. In particular, running all computers on the same platform would be highly beneficial both from an operations point of view and for the users at the DPINEs who can use the same technologies and thus share experiences (and meaningfully participate in the same courses). Budgetary constrains make this slow going with a time frame that at best covers all of 2005 and 2006.
- Ensure that the IT-infrastructure becomes tinker-proof. This requires a level of control that we currently do not possess.
- Ensure that each DPINE has two (or more) people with some basic IT operations knowledge. Some DPINEs already have these people while some have none. We should at least try to involve and educate those that already have a reasonable skill level.

5 Appendix 1. Terms of Reference

Task description and qualifications

Chief Statistical/IT Advisor

*** General responsibilities and tasks**

The Advisor shall assist in the capacity building in INE to undertake the tasks of co-ordinating the Mozambique National Statistical System, both internally in INE and at other producers and of supporting the statistical production systems. He/she shall collaborate with and assist the senior managers and other staff of the INE and other stakeholders, providing overall assistance, guidance, advice and training related to program-supported activities in order for INE to achieve the timely outputs of the Program.

A major part of the tasks of the Advisor shall be assistance in preparation and implementation of on-the-job training, organisation of formal courses and seminars, observation and participation in meetings (as ex-officio observer), field-visits to local governments, contact with other institutions working in the field of statistics and poverty monitoring, etc.

He/she shall also advise the institution in carrying out its mandated functions in general, and in particular support the capacity building in the horizontal functions both at the Headquarters, in the Provincial Offices and in the relation to the stakeholders.

The Advisor will report to the Team Leader, assist the Presidency and work closely together with the Directors of DARH and DICRE as well as with the Heads of the Departments in DARH and DICRE.

*** Specific tasks**

The specific tasks shall further include support to the following:

- Further development of the IT strategy and its implications on the long term planning including advice on suitable short term consultancies.
- Improvement of the overall management functions at all levels within INE, including the user side of IT, systems planning and strategic issues of co-operation.
- Implementation of an Administration and Management Information System.
- Human resource administration, with the aim to increase the efficiency and impact of training activities.
- The National Statistical System (SEN), i.e. to promote the data collection methods, to improve accessibility of existing statistics from the SEN, to improve the co-ordination and harmonisation of official statistics, and to strengthen the awareness and use of official statistics.
- The Internal Organisation of Statistical Information, improving accessibility and usability of survey data by using uniform methods for the storage of survey data
- Strategic matters related to IT, training of the subject matter specialist and improving the ability of INE to cover the needs of IT issues with own resources.

- The Provincial Offices, primarily aimed at developing and implementing improved data collection routines as well as other relevant duties to the 11 Provincial Offices.
- Building up general capacity within DISI in systems development and systems maintenance, and to improve the ways IT is used within INE.
- Training of non-IT staff throughout the organisation.
- Planning and implementation of Internet strategy
- Support the planning and implementation of project management

The tasks are not necessarily limited to the above.

* **Qualifications**

- A Master's degree in Economics Statistics, Business Administration or Computer Science.
- Minimum 10 years experience in planning and administration of statistical systems.
- Extensive practical, long-term experiences from methodologies within the area of IT and statistics.
- Previous experience from assignments in developing countries.
- Adaptability, social sensitivity and respect for a variety of cultures.
- Familiar with Scandinavian development assistance in general.
- Fluency in written and spoken English and an adequate working knowledge in Portuguese.

* **Working language**

Portuguese and English

* **Duty Station**

Instituto Nacional de Estatística (INE) Maputo with possible travel up-country to Provincial offices

6 Appendix 2. Scouting for resources at the home office

To learn more about the extent to which Statistics Denmark would be able to support us in addressing the tasks outlined in the above Inception Report, the present LTA went on a short-term mission to Statistics Denmark in order to scout for potential resources (December 17-23 2004). The present appendix summarises the main findings.

6.1 Summary of findings

During the mission the author met with a number of people to discuss the challenges facing INE. Besides strengthening the ties to Statistics Denmark, a number of current tasks and problems were discussed, and some relevant proposals came up. Notably, the data warehouse architecture was redesigned, and a number of issues related to the production pipeline, from the usage of the data warehouse through the output database to the publication on the portal, were discussed. Further, potential short term missions related to competency building for the support of the data warehouse effort were discussed.

6.2 Mission background

The Scandinavian statistics bureaus attempt to strengthen operations at Instituto Nacional de Estatística (INE), Mozambique, through the use of long and short term advisors (LTAs and STAs). Within the IT area, the LTA provides strategic as well as specific advice on problems in the area of IT and statistics production, and the LTA attempts to transfer knowledge specific to statistics production from the statistics bureaus of the Scandinavian countries. The current LTA has a background as an IT consultant and has only limited knowledge of the IT work at the home organisations. The mission, a visit to the home office, was undertaken to gain better knowledge of the IT side of Statistics Denmark as well as to strengthen social contacts with potential resources there. Also, the opportunity was taken to discuss the major IT problems currently facing INE in order to solicit feedback. Part of this feedback is outlined in the next section. Finally, attempts were made to identify areas where short term missions might be useful.

6.3 Findings

6.3.1 Data Warehousing at INE

In the draft architecture for the data warehouse of Panguana and Bormann we envisaged making aggregations on the fly in order to always provide users with fully updated information. However, the counter argument was presented that one should attempt to give users a more static perception of data. Thus, once statistical data are published they are the official truth, even if minor changes occur. In this view, error correction should only happen when major corrections occur, and it should be explicit. For this reason the published data should be kept in separate database, the output database, which should only be updated according to a strict (publishing) schedule. For

instance, a full update could be published every year while new data might be added on a monthly or quarterly basis.

The data kept in the output database should not only be aggregated but also analysed (using SPSS etc.) and annotated for users to be able to infer the quality of the data. It should, of course, only be published if the quality is above some threshold, but the methodology and processes/corrections for ensuring this should be readily available to the users. The aggregation process thus is not a simple one, and can only be automated in part, which further argues that the microdata should not be aggregated on the fly but instead the macrodata should reside in a separate output database of the same structure.

It was emphasised that the condition for any meaningful comparison of data (within the country, over time, or with other countries) and thus for sharing data with others, i.e., with users, is that metadata is available - and hopefully close to whatever international metadata standards that apply. (GDDS, IMF, OECD, etc.).

In line with this, the initial step suggested for the data warehouse implementation is to make a catalogue of data and metadata which lists, a.o: The data sources (survey, census, delegados, National Accounts, or other), the variables, and provide a code manual. Both at the micro and macro data level. Reaching agreement on the usage of terminology and codification across all departments of INE is foreseen to be difficult (but necessary) as the recodification initially will entail more work for the statistics departments.

The data entry module add-on to SPSS¹ should be excellent, and using it could potentially make integration with the data warehouse and the subsequent data aggregation (using SPSS) easier. It was even suggested that one could make a data warehouse solely using SPSS files, never introducing an actual, relational database.

6.3.2 The INE Portal

The coming data warehouse, and also the portal, places data in a common area at INE. It is then needed to grow a mentality, if not already present, where data and derived data are considered INE's data rather than the producer thinking of himself as owner of the data (even if he is responsible for, or owns, the process leading to the data). Thus, data should be readily usable to everyone. This requires that one establish a common rule set for usage of the data warehouse and the portal.

As one needs central control of updates, one cannot allow the statistical offices to define the rules of play at their discretion; these must be defined by a central authority. The statistics offices should still be the ones placing data on the portal, and be responsible for data maintenance, but there should be strict rules as to who does what, and where and when. Further, like with the output database, updates should happen according to a change request

¹ Currently CS Pro is used for a data entry module.

procedure, be scheduled, and be infrequent. (If desirable, one could have a 'news' section on the page).

For the more technical considerations, the concern was raised that file sharing or other personal downloads might be chipping away bandwidth and thus raising the load time of the web pages. Further concern was raised regarding the types of documents found on the portal, both because they are plenty and because some of them consume a lot of space and thus, if users are interested, a lot of bandwidth.

It was also suggested to do a thorough analysis of the website traffic in order to know what content is actually used. And use that information to clean up the website at a regular basis. Also, the PC-Axis database - and maybe even the output database - should only contain data that many people are looking for. (Infrequently used data and analysis can be created on request). One might also solicit responses from users of the portal as to whether they are able to find the information they are looking for and, if so, they find it easily accessible.

It was suggested that one ought to be running PX Web on its own server, not one shared with Plone. This should be done both to keep operations cleaner and to avoid contention for computational resources. Further, for trying to keep things separate, the implementation had PC Axis/PX Web use port 82 for communication. It was suggested that this might conflict with the security setup of some users, and thus might be the reason that some users experience very high latency when loading pages while others do not - simply because if port 82 is closed then PC Axis must start looking for another port over which to communicate. (An intermediate fix might be to have PX Web use port 80, even if one then cannot distinguish between Plone and PX Web traffic).

6.3.3 The migration of the INE server environment

It was suggested that the migration to Windows 2003 server be done separately from the upgrade of the Exchange server in order to make sure that the first migration actually works before touching the mail servers.

As for the migration itself it was suggested that one upgrade an NT 4 machine, for instance the backup domain server, in order to get a smooth transition from using a (Primary) Domain Controller to using Active Directory. The Active Directory should then be moved to a clean Windows 2003 machine (and the upgraded NT machine be removed) in order to get a clean installation.

Related to the migration the network/environment was discussed. Concern was raised as regards the fact that one still uses hubs, rather than switches, in most of the network. Further, the environment was considered 'thin' on emergency power supplies (UPSes). And the lack of an updated security solution was considered so worrying that an attempt to make a good deal with Trend Micro's Danish offices was made. (The effort succeeded, but only after DISI has bought new licenses - fortunately also at a considerable discount).

6.3.4 Potential missions

Three suggestions for short term missions, primarily with educational objectives, were conceived:

Structuring PX-Axis databases

People were impressed that INE had succeeded in getting the PC Axis/PX Web solution into production without further assistance than the initial pilot investigation. Still, they believe that the DISI portal group might benefit from learning still more about the solutions offering and further they would like to share their experiences in structuring PC Axis databases for increasing usability and easing maintenance.

Specifically, a hands-on workshop in PX Make which creates PC Axis files and is a good tool for structuring the PCAxis database was thought to be valuable. The focus of the workshop would be that what clusters of data one wants to present, and how, should drive the (re)structuring of the PC Axis database (which in some scenarios is identical to the output database), which again should be reflected in the structure of the output database, which again should be reflected in the structure of the data warehouse - and vice versa. Thus one could see a workshop in structuring the data of PC Axis/PX Web as an exercise helping to build competency for use for the data warehousing effort. (For presentation purposes it should be noted that PC Axis files also contain metadata, cf. the discussion of the data warehouse and the output database).

Database administration

According to plan, the INE data warehouse will be built on top of a (real, relational) database, but currently INE has only very little experience with such databases. Since relational databases are a central pillar of the IT environment at Statistics Denmark it was then suggested that one attempts to transfer some of the knowledge in that area.

Of particular concern would be disaster recovery from backup as this is tricky. Part of the problem is that one needs 'comparable' hardware as well as a complete list of software and patches and configuration changes etc². Further, one need full data descriptions to know what the data are referring to etc. This, as well as more general administrative tasks, and physical implementation of data models, thus might form the core of an educational mission

² For the data warehouse (micro data) and the output database (the corresponding macro data) having fail-over servers thus is a very appealing possibility.

Security

The knowledge and operational capability within the security area is severely limited. One might then use the staff at Statistics Denmark for giving a (mostly hands-on) course within this area. (One of the staff members there used to work for a security company).

I will add the proviso, though, that we actually know a good, local consultant who would be able to provide the same service. He has been used before whenever disaster was looming³.

6.4 Conclusion and Outlook

With the changes to the data warehouse architecture, the IT systems will be very similar to those of Statistics Denmark, if smaller and potentially cleaner, based on micro database (warehouse), and output database, and a PC Axis/PX Web front end. Thus we should be able to make good use of the experience already gained at Statistics Denmark as we proceed with the effort. The short term missions suggested all regard competency building/training, rather than gap filling, which bodes well for DISI's future ability to run the operations without much need for outside assistance.

6.5 Mission TOR

TERMS OF REFERENCE

for a short-term mission

on

Scouting for Resources at the Home Office

17 december – 23 december 2004

within the Scandinavian Assistance to Strengthen the Institutional Capacity of INE/Mozambique

Consultant: Karsten Bormann

Counterpart: Lars Erik Gevalli

Background

Before starting work as and long term advisor at INE, the present LTA within the IT area had the opportunity to interview selected personnel at Statistics Denmark regarding their jobs and, where relevant, about their short term missions at INE. Because the present LTA did not then know about the problems and opportunities facing INE, the discussion was very generic. Now the LTA knows a more about the specifics facing INE, but since his knowledge

³ The head of DISI says that she never is allocated funding for paying this consultant; he helps out because he is a friend of the staff. But asking him to make a course without payment is considered too much.

of Statistics Denmark is limited to what one learns during two weeks of interviews it hard to know where Statistics Denmark can add know-how to the operations at INE, or otherwise can be of use. Thus it will be of value to repeat the interviewing sessions from the point of view of the specific problems and opportunities facing INE.

Objective

To strengthen the relations between the LTA in IT and potential STAs at Statistics Denmark and to try to identify areas in which Statistics Denmark may be of use to INE through short term missions or other assistance

Expected results

If specific short term missions can be outlined, this will be considered good, but the main result is to increase the LTAs awareness of competencies within Statistics Denmark that are of relevance to the current situation at INE.

Activities

Initially, the LTA presents main observations during the first four months at INE to the people with whom further discussions are planned. These people will primarily be potential STAs.

During the following days the LTA will meet with these people individually to discuss the main points in order to get the potential STAs opinion as to ways that s/he would be able to assist. If during this discussion other potential resources are to surface, contact will be made to those people.

Preliminary program

Friday 17 December:

12.00 - 22.00: Participation in social event.

Monday 20 December

9.00-10.30: International Consulting (Lars Erik Gewalli)

10.30-12.00: Meeting with DSt project participants, Karsten Bormann presents observations followed by a plenum discussion (Invited participants: Lars Erik Gewalli, Anja Stiil, Niels Jespersen, Bo Guldager, Annegrete Wulff, Lars Knudsen, Jesper Ellemose, Søren Netterstrøm, Mogens Grosen Nielsen).
Meeting room 0029

13.00-16.00: IT Centre (Niels Jespersen, Bo Guldager)

Tuesday 21 December

10.00-16.00: Databanks (Annegrete Wulff, Lars Knudsen, Jesper Ellemose)

Wednesday 22 December

10.00-12.00: Meta data for the 'data warehouse' (Søren Netterstrøm)

13.00-16.00: to be agreed

Thursday 23 December

10.00-15.00: International Consulting (Anja Stiil)

Consultant and Counterpart

Consultant: Karsten Bormann
Counterpart: Lars Erik Gevalli

Timing of the mission

December 17-23 (5 workdays).

Report

The report will detail the main points presented to Statistics Denmark as well as suggestions as to issues where Statistics Denmark can be of use. It may also contain suggestions as to strategies that may be worth considering. Any troubleshooting taking place must be detailed elsewhere (but a reference can be given).

These Terms of Reference were prepared by

Day / /
.....

Approved by/in the name of the President of INE

Day / /

Prepared by: Karsten Bormann. Chief Advisor, Scanstat