STATISTICS DENMARK

Statistisk sentralbyrå
Statistics Norway

Statistiska centralbyrån
Statistics Sweden

Final Report from a long term consultancy on IT and Statistical Production

September 1, 2006

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

Karsten Bormann

_____
Instituto Nacional de Estatística

September 2006

Karsten Bormann
Instituto Nacional de Estatistica
Av. Ahmed Sekou Touré 21, 5
Caixa Postal 493
Maputo
Mozambique
karsten.bormann@ine.gov.mz
++258 84 56 86 88 0
or
moz@bormann.ac
++45 43 54 07 03

Table of contents

1 Executive Summary

The IT infrastructure has expanded rapidly during the last several years and has reached the consolidation phase. The exception is the support for statistical production offered by the data warehouse effort, which is only in the prototype construction phase.

DISI is in a rather healthy state: Its management has good grasp of the fundamentals of IT management and technicians are reasonably capable, though capabilities and commitment vary a lot within this group. Importantly, those servicing the basic infrastructure are qualified to do so.

Still, execution is nowhere near smooth. The main reasons for this can be traced to

- Lack of access to money for executing any given task, maintenance included. This is true even when the activity was budgeted.
- Lack of access by DISI management to its own staff. Further, staff is often used by others without asking permission and without giving notification.
- Lack of skill and commitment by certain staff, exacerbated by DISI's lack of control of its own personnel.

These root causes of the execution problems are largely outside the influence of DISI which suggest that, to be effectively addressed, one needs to change the general management procedures surrounding the use of IT resources and IT personnel. Specifically, two main approaches suggest themselves. The first is for top management to take a more direct interest in the IT area, including execution. The second is to hand over the authority to make financial and staff decisions to the management of DISI. Ideally, the two approaches would be combined.

I would have liked to take part in such an effort to address management issues more broadly, but all good things must come to an end. It has been a privilege to be here.

2 Introduction: Birds-eye view of the IT infrastructure

The basic IT infrastructure at INE consists of a wired network at the central office and wireless networks at provincial offices. All networks contain, among other things, a file server where data and documents can be stored and systematically backed up. All employees have access to the network (email, internet, etc.), though some employees in none core functions share computers.

Micro data from surveys is managed by the IT department, and is mostly stored in CS Pro files, Access databases, and Excel spreadsheets. Macro data, or aggregated or analysed data, is stored on employees' workstations and on the file servers. Most new published data also reside in a PC Axis database accessible from the INE Portal.

Data entry applications and data storage applications are developed, mostly in CS Pro or MS-Access, by the IT department which, furthermore, does most of the tabulation within the area of social statistics, recently using SPSS.

The following sections outline the current state of activities related to the most important parts of the IT infrastructure as well as to the supporting management structure. Recommendations and suggestions are given along the way and are summarised in Appendices A and B. The discussion outlines the views of the author which do not necessarily coincide with the views of Statistics Denmark, Danida, or INE.

3 The basic infrastructure: INE and DPINE Networks

3.1 Principles

The overriding principle for infrastructure management is that it should be kept simple to be kept efficient. This means homogenizing the infrastructure, using the same software for the same task - and using as few different components and software as possible.

On the technical side, this is done to minimize problems with interoperability. On the IT administration side, it is done to make it easier for the system administrators to know the technologies involved. This minimizes the number of personnel that one needs to service the infrastructure as well as their training costs. It also eases troubleshooting in case of problems, as the systems are simpler.

On the usability side, using homogeneous systems increases the likelihood that a nearby colleague knows the solution to a problem. And it increases flexibility because the tools are the same throughout the organization for which reason people may more easily cooperate with others or move to different departments all together.

These basic principles now feature prominently in the thinking of DISI.

While all this is very important at INE itself, it is extremely important at the DPINE because they use the INE technicians, separated from them by a day of travel - as well as by travel costs, the funding of which is extremely slow going. This in turn effectively makes the IT department unable to deliver a reasonable maintenance service level at the DPINE.

In my opinion, technological solutions are generally being oversold. Having a sea of features, they are able to do everything. Absolutely everything!! It is like a flashy Mercedes, beautiful to look at and good for showing off. But if one really wants to go places, one is better off with a 1970s Land Rover Defender. It can go almost anywhere and it is extremely simple, so it can be repaired everywhere. We must make sure that our IT operations are as simple[1] as the Land Rover. We might not feel as good next to the flashy Mercedes', but we will get farther.

---

[1] This is known, colloqually, as the KISS Principle (keep it simple, stupid - or 'sweetie', in these modern days) and probably is the most important principle of software development and IT configuration ever formulated.

3.2   State and Future of the INE Servers

The INE IT infrastructure is in reasonably good shape, not least because of the permanent presence of the technicians and because funding often is available at the statistical departments. Still, an overhaul of the central servers was long overdue as one has experienced rather frequent breakdowns and also the operating system of part of the infrastructure (Windows NT) no longer gets updated when (security) bugs are discovered. And the options for forcing workstations to adhere to the standards set in the IT Policy, thus simplifying the operations, were not the best.

For this reason, a migration to Windows 2003 has just taken place (August 2006). It is instructive to review the history of the migration. First time it was scheduled to occur was in August 2004, but securing funding for the migration proved hard work. Still, in January 2005, training in migrating to, and using, the new environment has been completed. Migration procedures have been planned and thoroughly tested. And recovery procedures for the new environment have been tested and documented. (This was done both by technicians themselves, after the course, in 2005, and during a short term mission in November 2006). The actual migration only takes place as I am writing this, the delays being due to the slow progress on securing its funding and actually having the money made available. The migration should leave INE with a rather stable IT infrastructure for the next three years or so. This time can then suitably be used to carry the current discussions on platform and architecture to an end and to do the needed training and testing in case of a change of platform.

The slow progress on improving the IT infrastructure - both at INE and DPINE - belie the fact that DISI has qualitatively improved and is in reasonably good shape to make the needed changes happen. Before, consultants made migrations and technicians were uncomfortable making changes to this initial configuration in case it would bring the installation down. Now, technicians know that it is *their* responsibility, feel comfortable about it, and - even if the actual migration took a long time to happen - are well capable of migrating the servers and operating them afterwards.

3.3   State and Future of the DPINE Networks

One should be positively surprised that the IT environment at the provinces, or the DPINE, is at least partly functional. Being, as they are, networks of approximately 10 computers ranging from new to obsolete, running many different operating systems (all Windows), containing too many printers, and all connected with 'hobby' network equipment, and linked to the internet by a modem. Not to mention the support technicians being up to 2000km away with next to no resources allocated for maintenance. There seems to be a wide gap between the requirements to the DPINE IT environment and the willingness to support it financially and institutionally.

However, once the level of ambition has been chosen, the requirements regarding the implementation of the infrastructure are rather rigid. For the

DPINE, the current policy is to provide all users with email and Internet access. This requires a functional internal network, based on professional equipment, and it requires a fixed digital line for Internet access (which in turn is a prerequisite for email access). If one is not prepared to pay what this costs[2], one is better off disabling the internal network (use memory sticks for file transfer) and only connect the workstation of the delegado, via modem, to the Internet.

Following the current strategy, a model environment has been created in Inhambane (and partially in Gaza), even if based on low-end equipment. The training effort is taken seriously and How-To documentation has been made to support DPINE users. The need to simplify and harmonize the user interfaces is also well understood and implemented when opportunity occurs. And, when people of the right qualifications are present at the DPINE, we teach them to make simple computer administration tasks. In short, given the opportunity DISI should be capable of securing the DPINE IT infrastructure. The opportunity will only occur, however, if management support of the effort becomes more pronounced than has yet been the case.

The lack of management support for improving the DPINE infrastructure is illustrated by the case story in Appendix C (section 11).

---

[2] With 10 to 20 users, using a modem for consistent internet and email access may well prove a more expensive than connecting via a fixed digital line with flat-rate payment.

4 Infrastructure for Statistical production: The Statistical Data Warehouse

Once the infrastructure has been taken care of, it is time to consider how to support the production pipeline systematically. To this end, we have elaborated on the notion of a Statistical Data Warehouse. It is now well thought through, and partially worked through. And the actual implementation chosen has been modified to reflect both the level of professionalism of the technicians and the level of control (/access) that DISI has over (/to) its own personnel. The Data Warehouse is by now a very simple conception, but it has kept all the key features of a statistical data warehouse. It probably is important to see the Data Warehouse not as a piece of technology but as a support of the production pipeline at key stages. Thus, whatever we do, we should embed in the solution the ways we ideally would like to produce statistics

4.1 Purpose and Function

Before describing the architecture, it is worth reviewing the support of the statistical production pipeline expected from this initiative. The Statistical Data Warehouse consists of three parts

- Micro Data Warehouse - For keeping the cleaned micro data and the data received from the ODINE
- Macro Data Warehouse - For keeping analyzed or aggregated data and ODINE data in the (perhaps modified) form they will be used by INE
- Dissemination Database - For keeping data that is available to all users

Obviously, these three parts of the Data Warehouse are to safeguard data at three different stages of the production pipeline and make it available for further analysis. (In the case of survey data, one first has a cleaning step during which data is stored in CS Pro files). That is, the role of the Data Warehouse is to support the *statistical production pipeline*. Conversely, for the Data Warehouse to live up to its potential it eventually needs to be driven from the statistical departments.

It is important to note that even if one did not call it a 'Data Warehouse', and did not implement it physically, one would still need to implement it logically if one is to secure the quality of the statistical production. Because one needs to know:

- What micro data one has
- What data one publishes
- What (common) classifications one is using[3]
- Where data is stored

Keeping logically track of all this information is much easier if one has a physical structure that corresponds to the requirements.

---

[3] When the same statistical variable is published using different classifications from data source to data source (as it for instance happens in ESDEM) one loose the opportunity to create good time series of data and thus greatly reduces the value of data.

4.2 The Architecture

4.2.1 Micro Data Warehouse

In the current conception, survey data will be digitized and cleaned using CS Pro while ODINE data may arrive in whatever format, even on paper, and will be kept or captured in its original format. This micro data (and raw data) will be stored in the Micro Data Warehouse. The Micro Data Warehouse is as simple as it is flexible. It has a Data file for each class of survey objects, such as *Person* or (also known as a fact table) and for each such survey object it has a Classification file containing the value sets (also known as dimensional tables) pertaining to each variable in the data file. All data are stored in 'comma separated value' (.csv) files, in Ascii format, using only characters A-Z, a-z, 0-9 and _ (underscore) to assure that it can be read and used - now and in the future - by more or less all software. Metadata such as a .tiff image capture of the questionnaire, reports on procedures, analysis etc is to be kept alongside these two base files. Documents preferably should be in XML tagged text files, perhaps along with a .tiff image capture or, if one allows for less robust formats; a .pdf capture. (Most formats different from text can be though of as lost to the historians).

It is worth noting that, while simple, this is a complete implementation of the star model (or multidimensional modelling) used in traditional data warehousing - with the added metadata focus of statistical bureaus.

If one day needed for performance, the .csv files can easily be imported into whatever relational database[4] (and whatever analysis tool). But the fundamental Data Warehouse is a simple directory structure keeping data and metadata in a robust, non-volatile, and easily exchangeable format. Even if one does introduce a 'real' relational database one will need the same structure *and* one will need to keep data in the robust format in the directory structure *as well as* in the database as an archiving effort. Otherwise one will run the risk of loosing (all) one's data.

In the current conception, the data and metadata will be reflected into the World Bank's Metadata Toolkit. This is a nice framework as it makes it obvious when metadata is missing and thus may be able to support the quality of our statistical products. However, because data is kept in non-human readable files, the same caution exists as for a relational database: The data **must** also be kept in an analogous directory structure in simple human readable formats in order not to imperil our entire collection of statistical data. It thus is easiest to think of the directory structure as the fundamental Micro Data Warehouse and the Meta Data Toolkit or a relational database as tool for presentation.

---

[4] The current amounts of data does not justify this, and the shortage of (qualified) staff at DISI suggests that it is presently not a very good idea. The more complex (and expensive) the database, the worse the idea. If one introduces a relational database, it should be for staff rentention purposes, as the staff thinks relational databases are more sexy.

4.2.2 Dissemination Database

For the Dissemination Database, we are already using a PC Axis database. Here, data is stored in Ascii based text files, i.e., the format is immediately human readable with all current text editors. And, again, introducing a relational database seems to add little value but lots of fragility. Thus, it was decided to stick to PC Axis.

Another contender was ESDEM/DevInfo, but it is only able to classify statistical variables according to area and subpopulation (and time), which is too restrictive. Practically, people often want to classify a statistical variable according to more than one dimension, say; gender, age and urban/rural. In ESDEM/DevInfo this can only be done by allowing a proliferation of subpopulations - as can also be seen in the data published in ESDEM. And this imperils both data usage and classification quality, as can also be seen in ESDEM. But the policy oriented front end and data browsing capabilities of ESDEM are very nice, and we have tried to enable this type of structure in a prototype based on the PC Axis database.

4.2.3 Macro Data Warehouse

The relationship between the Macro Data Warehouse and the Dissemination Database is that the Macro DW stores the end results of data analysis, whether it is published or not, while the Dissemination DB stores analysed data that meets requirements of anonymity and is deemed of relevance to a sufficiently large audience. The very easiest way to move data from the Macro DW to the Dissemination DB is if the two have exactly the same structure and use exactly the same technology (i.e., currently PC Axis). Then one can simply tag things to be published along with a date in the Macro DW and, on that date, copy the data to the Dissemination DB. This construction can also be used to build user acceptance:

Currently, DISI (informally) is the keeper of micro data and also make a fair part of the tables within the area of social statistics. Thus the Micro Data Warehouse primarily will be a help to DISI in making data more easily available for analysis. (Although, eventually everybody should notice the increased transparency of data and metadata). The Macro Data Warehouse, on the other hand, should function more or less without DISI involvement as, in principle; DISI is not doing the statistical analysis. Thus the question occurs as how to get the statistical departments to take up the burden of storing data in a systematically way. The easiest approach seems to be giving the statisticians the responsibility of storing their own data on the Portal, i.e., in the Dissemination Database. (One should make sure it has the desired structure before doing this). Then, when they have gotten used to publishing data on the web, and feel ownership for that process, one simply creates the Macro Data Warehouse by cloning the Dissemination Database. After this, statistical departments will store their data not on the web but in the Macro DW, and at the publishing date it will automatically be pushed to the Dissemination DB. Thereafter, one can store things in the Macro DW even if they are never going to be published (important partial or intermediate results, non-anonymous data, data with little interest to the general public, etc.). That is, it is suggested to build user (producer) acceptance by building from the Portal.

4.3 Increasing quality and utility of the Data Warehouse

When discussing the statistical production one often talks a lot about the users of the products. However, we do not really know our users. Further, we lack a large research community (as well as the resources for interacting with it) to drive requirements to, and thus improvements of, our statistical production. Lastly, we do not do much analysis in-house. This means that we lack drivers for improving quality and content of the Data Warehouse. One needs to compensate for this. Probably, having sophisticated users in-house would create the most impact. Thus it is suggested to create an analysis unit in-house to drive the development of statistical quality, including that of the Data Warehouse. This might consist of the handful or so persons who were educated in statistics as well as perhaps a few other staff. (If not possible on a full time basis, then at least half the workweek of these people should be with analysis). Given the opportunity, adding analysts from other parts of the national statistics system (SEN) to this group probably would beneficial, as better integration of the (dissemination of the) data produced by SEN could boost the availability and utility the data. Or one could even join forces with some of the independent statistical analysts working in Maputo, such as those of the World Food Programme, or Arnt Channing at the Ministry of Planning and Finance.

Finally, one should think about INE's role in other peoples' production pipeline. It has been suggested to try to sell data or publications on the Internet. Besides having turned out not to be worth the effort in other places, it seems a bad idea to withhold statistical information of general relevance from the very public that funded its production. It seems much more promising to supply consulting services, initially in the form of information gathering, later perhaps augmented by analysis. Why would people pay for such a service when (most of) the data is freely available on the Internet? Because their own employees would spend much more time locating the relevant data than someone who is familiar with the data, and they may end up not reaching all the relevant information. Gaining the level of control of our data that is the ambition of the Data Warehouse effort would make such a business model rather more attractive.

5   Process and Project Support

Having outlined the strategically most important DISI project, it is well worth spending a little time on discussing what it takes to make such projects succeed.

5.1   Basic Management Information

In order to make project and people management more efficient it is at the very least necessary to have a list of approved projects, tasks, and subtasks and have the staff register their time usage according to the task carried out. One need to know how much time goes into a given task or product in order to decide whether carrying it out is worthwhile. And one needs to know the time spent to calculate performance on different task types in order to do predictive scheduling (or planning) of future tasks. In cases with large differences in performance between staff members on like tasks one also can have the high performers teach their peers to become more efficient. Of course, one would also make visible the performance of employees - which might be a motivating factor in itself, especially to the occasional bad egg in the basket.

To enable this, a (very) simple management information system (SIMIS) consisting of just three spreadsheet templates was created. It was never employed, though, in part because the experiences with the introduction of a quality assurance system showed that for DISI to go alone in changing work management in fundamental ways is difficult at best. Attitudes have moved a bit - some people are now creating some documentation of solutions and procedures - but for things like time registration an institution-wide pressure needs to be applied. The other reason for not introducing time registration without this being implemented at all of INE is that DISI staff is used by others (often without the knowledge of DISI management, see later) and thus these third parties would have to use the management information system as well, in order for it to make any sense.

5.2   Quality Assurance and Quality Management

As mentioned, the quality of the work processes has been under scrutiny and - for the data modelling related to the Data Warehouse - a pilot effort was made to introduce CMM (Capability Maturity Model) inspired quality assurance techniques, such as systematic reviews, version control, and acceptance procedures. It is would be an understatement to say that the average technician has a hard time seeing the use of managing quality. Again, the impression is that unless this is driven from the top, giving the procedures a sense of finality, such measures are not going to deliver on their promise. In the case of the Data Warehouse we have therefore opted for gaining acceptance the slow way, through the back entrance, by taking the tools already in use as the starting point and have people modify the products that were already made - for instance modify variable names and value set descriptions in order to create usable documentation and datasets. Hopefully, then, once the documentation starts showing its worth - showing that proper procedures is not only a benefit to management but also to the technicians - it will motivate the staff to do a more systematic effort in this area.

Note in passing that a making systematic effort in this area is as important for the statistics departments as it is for the IT department, and that one preferably should build a common system rather than two or more separate efforts.

5.3 Flexibility and Robustness (the 3x3 Rule)

To make execution itself more robust, the IT personnel has been grouped into three groups, each of 3-5 employees: Infrastructure administration, application development, and application maintenance and servicing, and training plans have been made to harden these groups. By consolidating on a common set of tools for all tasks carried out by these groups, we are trying to get to the situation where each employee within a group is in principle able to carry out all tasks of that group, increasing both robustness and flexibility of the execution. This is often known as the 3x3 rule; all main tasks should be covered by at least three staffers, and all staffers should be able to cover at least three main tasks.

The groups should be big enough for the members to have sufficient problem solving skills between them and small enough for the members to be able to share knowledge efficiently. Also, the groups should in principle be self-managing (the administration group already is) so that they for most purposes have the responsibility for execution and the authority to choose the technical side of solutions.

In spite of this emphasis on managing employees by the group, however, we have found that vast differences in employee motivation and work ethics in most cases compel us to attribute tasks to specific persons. Further, while the infrastructure group is in rather good shape, the rest of the effort is not going well: There is as yet no funding for training, and DISI has so little control over its own personnel that meaningful reorganization of work practices might be a guiding principle when opportunity strikes but is out of reach as a systematic effort.

6 Management support

The most severe problems at DISI occur because the head of DISI (and the adjoined director) has little control over the budget execution and the staff. The present section discusses the associated problems and how support from the top management could change the rules of play.

6.1 Popular DISI

When blame games are being played at INE, DISI is very often at the receiving end, blamed for delivering insufficient service of insufficient quality. This is strangely at odds with the fact that other departments and directorates are extremely eager to use DISI personnel, often for tasks one should think would be solved by the statistical departments themselves. With half of the DISI staffers so popular that other directorates and departments want to grab them at every possible time - often without asking, or even notifying, the head of the department - one should safely be able to pronounce DISI the most popular department at INE, and the blame games primarily a way diverting attention from oneself. After all, it is easier to blame one's problems on the IT department than it is to assume responsibility oneself.

Another thing is that different staff members *do* have different skill levels and - as importantly - different responsibility levels. And while some deserve credit for their good and dedicated work, others may not.

It seems a weakness of the system that there is virtually no link between performance and compensation and, related, there is no such thing as a technical career track. Often people are paid vastly different salaries for doing exactly the same job (due to differences in educational background). It is important to make a concerted effort to motivate the staff, especially those lower than 'head of department'. Besides financial rewards (or threats) one has possibilities like; make sure people develop their skills base, make them proud of their job by communicating the importance of their deliveries, and make sure to praise them whenever a job is well done. Too little praise (and too much blame) seems to be bestowed on our technicians. (That goes for the DPINE, too, and maybe also for the rest of INE?).

In the case of unmotivated employees (the bad eggs in the basket), the lack of a good, simple management information system is being felt particularly hard as there is no good overview of the tasks attributed, the time spent on their tasks, or the performance on those tasks. When people work at more than one department, as is the case for the DISI staff, one cannot know if a person is absent due to work or due to pleasure. Only if there always existed clear-cut agreements about other departments' use of DISI personnel (an agreed task, for a specified person, during a specified time interval) could one split of the management task related to the shared employees, and thus give praise where praise is due and blame where blame is due. Currently, such a key ingredient of a systematic approach to personnel management seems the exception rather than the rule.

It is a severe constraint on DISI's ability to deliver in any strategic development effort that its best staff members are grabbed by others all the time - without them having to wait for these people to complete their tasks at

DISI, and often without going through the head of department. At DISI, this situation turns planning into wishing as the head of the department has no control over where the staff will be employed, by whom, and when. Thus, any project spanning more than three man months with subtasks spanning more than a week is not likely to get off the ground and one will have to set the level of ambition according[5].

6.2   Money constraints

Another severe constraint, hampering smooth execution, is the lack of budget - or even of *access* to DISI's budget, after it has been approved. As effectively demonstrated by both the processes surrounding the migration of the INE Servers and the overhaul of the DPINE environments, DISI can only meaningfully plan and work on things that do not require any sort of financial input. Whenever money is needed, initiatives come to a complete standstill. Even when budgeted, the administrative work involved in getting access to even trivial amounts of money is excessive. For an IT department, this is a showstopper.

This situation might be partially remedied by switching to Open Source software and by letting the DPINE take care of their own IT environments (which they would then need to secure funding for). But the problem is likely to continue keeping DISI from living up to its potential. And, as said, the problem is not just about money but also about the budgeted money actually being made available when it is needed for execution. Even considering the limited resources available, we could have done better.

6.3   Is execution as good as it gets?

DISI is technically reasonably capable, and there is good understanding of what it takes to run the department successfully. But circumstances often do not permit this. Thus there is the prospect that this is as good as it gets: The factors on which potential improvement of the execution at DISI rests are largely outside (the influence) of DISI - and it may well be that it is not possible to make DISI qualitatively better without changing the way INE works. Or adding significantly more resources. Or both. At least, the most visible technical problems have institutional root causes as we fail to adjust the level of ambition to the resources present, fail to secure financial and human resources (and make them available), and generally do not stick to a course of action once it has been agreed. Etc. With real support and follow-up from upper management, most of these problems probably would go away rather soon.

One way of doing progress in this respect is to always have an activity accompanied by an agreed budget. This money must be (easily) available to those responsible for execution in order to avoid disrupting the execution. At DISI this probably would mean the head of department. One needs an audit trail for making sure funds are not misappropriated but, once the plan has

---

[5] The Data Warehouse effort was, partially for this reason, significantly reduced on the technical side. But by making the implementation technically simpler it was still possible to maintain the key characteristics of a Statistical Data Warehouse, and thus also the benefits implied by its implementation.

been accepted, the head of IT should need no further authorization to carry it through. Regular reporting on progress and budget execution is to be required, but not authorization before continuing the implementation.

A complementary way to progress is to increase information flows and to involve all layers of the hierarchy in important tasks. As a minimum, one could resume the monthly meetings between presidency, directors, and LTAs - with heads of departments called upon as necessary. To exemplify: In the case of the DPINE IT,

This group could have decided on the service level needed. DISI could have calculated the budget needed for delivering this level of service. Subsequently, one would expect some juggling until the service level matched the resources one was willing to allocate. And, when first the project had been put in motion - with an agreement as to its level of importance - this group would be able to make sure that the allocated technicians were working on that task, and none other than that task. Further, they could follow both progress and budget execution and help with removing the occasional roadblock. In short, INE has lots of competent employees and they do a lot of work, but with more communication - within and between layers in the organization - INE could still improve execution a lot, even on the current resource base.

If one relies solely on the latter approach (and maybe also if one chose the first), the IT area should have its own director - with no other responsibilities - in order to assure that IT and its usage is the sole success criteria for this person and that securing its place in the organization, resources included, is the way of achieving success. This would also allow the head of DISI to spend more time as a project leader and/or tech lead and less on paperwork - thus being able to follow up more closely on projects. The administrative side of running DISI (budgets, strategy plans etc.) would then be the responsibility of the director. In case one is not able to give IT its own dedicated director, it might be advantageous to either split up DISI or place it under the auspices of a director that has urgent needs for a well functioning IT infrastructure. If the provinces weight in heaviest, this probably should be DCI, while if statistical analysis weight in heaviest, it might be DEMOVIS. Because the DPINE is the most fragile 'part of' INE, and since money tends to follow the large surveys, I would opt for DCI. It is my belief that if DISI stays within DICRE, and with no financial authority bestowed on the head of the department, DISI is now as good as it is ever likely to get.

7   Epiloque

Many are the problems, within DISI and facing DISI. Such problems belie the fact that DISI has become stronger. A lot stronger. DISI more or less knows what is good for the IT infrastructure, for the creation and maintenance of applications, and for the management of IT resources. As the STAs from Statistics Denmark tell it: In the beginning every proposal was accepted without much discussion, now they get told when things do not make sense according to the actual situation or the plans for the future. They appreciate this. I hope this service extends to both the presidency and the directors.

INE is an ambitious institution. This ensures a rather large output which in turn often leaves too little time for building skills - be they individual or

institutional. That said, my overall impression is that the majority of technicians and most of those of higher rank take pride in contributing to creating good statistics and that they have skill levels that enable them to do so. This still leaves us with the task of motivating and training the rest, of course, but it is certainly a good basis for achieving results in the future.

All in all, it has been a pleasure to work at INE. Things have not moved as fast as I initially expected. In hindsight, this is no surprise because one is not only changing procedures but also mindsets, and this takes time. Still, things have moved. And with good colleagues and an ambitious organization, it has been both pleasant and interesting to be part of INE.

8   References

I have authored a number of strategic documents during my time at INE. From those, I have chosen those that seems of future relevance to INE and to DISI and placed them in a document repository. I also have included a number of documents by my predecessor, Mogens Grosen Nielsen. This repository can be found at:
DISI \ bormann (on which server it resides, I do not know, due to the migration, I lost trace of it ;-)

Below, I outline which documents can be found in the repository.

Reading Guidelines for
Document Repository - IT Long Term Advisors - Final
August 17, 2006.

The most relevant have been denoted **key document** or **important**.

The **root** directory contains one document:
- DISI Problems - perhaps the **most important** of all
- kb Final Report - the present document.

Then there are a number of directories of documents, the documents of which will be introduced in order.

Data Warehouse
- Data Warehouse Implementation - Shows some of the tasks envisioned after choosing an minimalist approach to data warehouse creation. **Important.**
- DISI News - Data Warehouse - The **key document** at the present moment. Discusses the architecture and the reasons behind, and outline the minimalist approach taken.
- INE Data Warehouse - Soeren Netterstroem's report on Data Modelling and the generic architecture of a Statistical Data Warehouse

DPINE - Provincial Email

- DISI News - DPINE - A **key document** holding my most recent writing on what we would like to do with the DPINE. It details: Current problems, possible solutions, and the solution advanced by DISI.
- DPINE Inhambane - 29.6.05-1.7.05 - Describes the mission to homogenise the architecture of the Inhambane office as well as problems remaining. The report is in Portuguese, the notes attached at the end are in English.
- DPINE Outsourcing Requirements - A **key document** (draft) outlining the services needed in case one wants to outsource the maintenance of the DPINE.

Management
- External Review Comment - The review comments I actually provided to the external review board. **Very important**.
- Evaluation Thoughts - My ideas of what needed to be communicated to the external review board. Probably important.
- DISI IT Budget Needs 2005 - For Steering Com June 2005 - Was an attempt to secure the budget for the most urgent priorities of INE, according to the management. That is, the DPINE communication and consolidating (migrating) the INE server environment. Worth viewing for it's discussion of the problems, the suggested solutions, and the link to funding. And as a reminder that funding, or lack of access to funding, has paralysed DISI for a long time - with the management (and donors) well within the information loop.

People Management
- DISI Education Strategy - An attempt to make a comprehensive plan to update staff skills in areas where we are week.
- INE IT Management - SADC Meeting.Febr.2006.doc - A short file showing how we envision DISI split in autonomous groups within the areas (3x3 rule); System Administration, Application Development, and Application Maintenance/User Support. **Important**.
- SIMIS - SImple Management Information System - A **key document** detailing the smallest possible information gathering that could significantly boost management capability. Not used. (Associated spreadsheets are in the QA Repository).
- Joel On Software - Painless Software Schedules.doc - A highly recommended background document to the SIMIS system/document.

Quality Assurance
The root contains Anastácia´s note on the processes for document creation. There are three subdirectories: *Manuais - Mogens Grosen Nielsen, Templates - Bormann, and Two Ideas For QA Directory Structures.* These will be outlined in turn

Manuais - Mogens Grosen Nielsen
Here, I have stored what seems to be the final versions of the Manuals made by Mogens and DISI as part of their work on building a quality system. These mostly contains outlines of procedures. The documents contained are:
- Desenho Especificações.doc
- Manual desenvolvimento ver12.doc
- Manual desenvolvimento ver21.doc
- Manual documentaçao VER21.doc
- Requisitos especificaçõesIFTRABv2.doc

Templates - Bormann

Which contains the templates I created to actually start using quality assurance practically. It contains three subdirectories, the files of which will be listed below:

Documentos e Revisão

- InqueritoNome - MLD do QuestionárioNome - Template - A3.doc -- Use this when making a Logical Data Model
- Revisão - Documento - Versão - Critico - Data -- Template v. A3.doc -- Use this for soliciting response to a document written. Should be stored along with the original document in order to document actions taken.
- Aprovação - Documento - Versão - Data -- Template v. A3.doc -- More or less equal to the above, but for the final version. Important documents should be formally approved to document that a task has been completed.

Guias

- Estimativo do Horário - Guia - A2.doc -- An attempt to boil down and translate the method outlined in the document ' Joel On Software - Painless Software Schedules.doc'. Used with the template 'Estimativo do Horário - Template - A1.xls'.
- Metodologia - MLD - A1 - CriticoBormann.18.10.05.doc -- Draft - Anastácia's attempt to write down explicitly the approach to creating Logical Data Models (Using the template 'InqueritoNome - MLD do QuestionárioNome - Template - A3.doc').
- Programação padronisada - Coding Standard - Guia - A1.doc -- Naming rules to be obeyed when creating Logical Data Models. One is better of using more or less the same rules when programming software, as the rules are meant for compatibility with the rest of the IT world .

Tempo e Estimativos

- Registração das Horas - Template - A3.xls -- Time registration on specific, well defined tasks is the basis for all real planning, i.e., for producing estimates rather than guesstimates.
- Desempenho - MLD dos Inqueritos - A1.xls -- When time use on tasks is known, one can calculate the speed at which given tasks can be completed.
- Estimativo do Horário - Template - A1.xls -- When the speed of implementation is known, one can make real estimates of planned activities. One can also use the personnel's own idea of how long a task will take. And train them to become better and better at estimating time usage. This template use the ideas of 'Estimativo do Horário - Guia - A2.doc' and ' Joel On Software -Painless Software Schedules.doc' to accomplish this.

Two Ideas For QA Directory Structures

Contains two essentially empty directory structures:

Data Warehouse - Repositório

Contains a preliminary structure for the directories in which all documents of the Data Warehouse should reside. Only the Micro Data Warehouse has the content indicated. In a sense, **this is the Data Warehouse**.

QA Repositório

Contains the structure in which Quality Assurance documents ideally would be stored. There are hidden a few documents within this structure. Primarily, there are a number of templates relating to QA in the subdirectories of QA - Repositório\DISI\Templates

Also, there exist a file related estimating projects in another directory, namely:

QA - Repositório\DISI\Info\Gestão de projectos\ Joel On Software - Painless Software Schedules.doc.

Strategy

- Initial Findings - Much of this is outdated, but much is not. Still **important** are sections 1, 4, 6, 7, and 10. Partially relevant are sections: 3.1, 3.2, 3.4.
- Vision - A **key document**. Short, omitting many details, but it gives an overview of the main parts of the vision for the IT environment (a year ago).
- DISI News - Architectures and Technology Choices -- This document is **important** for revisiting the choice of infrastructure platform. Discusses the present architecture, Open Source software, and an architecture based on a server cluster servicing flat clients.

9 Appendix A: DISI Problems and Solutions

DISI Problems	Description	Solutions
1. Lack of (access to) money, even for activities planned in the budget.	1. The overriding problem at DISI is the fact that every time money is needed, we are brought to a complete stand-still. 2. This is not just a matter of the budget being too small. Even when a task is budgeted, the administrative work involved in getting access to even trivial amounts of money is excessive. In fact, the planning and budgeting effort seems to little avail as it is no easier to get money freed for planned, budgeted activities than it is for ad hoc urgencies.	1. Place the responsibility for money where the responsibility for execution is, i.e., presently, with the head of department.
1. Lack of budget /money /access to money for equipment, training, current expenses	1. Without a certain, stable commitment in terms of budget - and strategies that are in harmony with the amount of money set aside for running the infrastructure (or vice versa) - planning becomes next to impossible. As does creating solutions to help reduce the problems that DISI does experience in execution. 2. A lack of knowledge by non-DISI personnel of the budget constraints upon the work of DISI makes the problem even harder felt. 3. This tends to demotivate even good staffers because they know in advance that most initiatives bestowed on them will lead nowhere.	1. We need to determine the long-term budget available and create a strategy accordingly. This may or may not include: Fewer Computers/users Fewer/simpler software Less DPINE support Open Source software 2. Avoid requirements/requests for which there obviously is no money. Otherwise one demotivates the technicians. 3. Have requests accompanied by the money to execute.
1. Lack of control of activities and personnel	1. Other departments and directorates often do not like to specify the tasks they want done, and the time it is supposed to take. This increases contention over scarce (human) resources, and make it DISI's problem when third parties want to use more resources than are available. 2. Often, tasks are then attributed directly to technicians, without the involvement of the head of department (or the director-ajunto). This jeopardises any planning capacity we might have as we loose track of what work is (or should be) done. It also makes impossible any attempt to prioritise tasks - and what resources to use for a given task - which in turn also destroys the planning capacity. 3. When tasks are not owned by DISI, but by other directorates and individual technicians, one cannot impose quality assurance on tasks and processes. 4. And finally, it removes our control over staff. We do not know whether	1. Better planning in other departments. 2. A more humble approach to the use of DISI resources, taking them in as equal partners. 3. Attribute tasks only through the head of department. 4. Physically have all technicians work at DISI unless the task requires them to be elsewhere. 5. Have a large whiteboard with the schedule for the week of all employees (including *where* to find the employee while working on that task). 6. Try to improve staff motivation.

	people not present at DISI are working elsewhere or are in town for the day. It depends totally on the self-discipline of the employee. And while some are very good, others may have less high standards.	1. Consciously make sure that good workers get attention, praise, education, vacation, support. 2. Attempt to make a career path for good workers, i.e, financial award or upward mobility. One way might be by having not only a 'management' based career track, but also a specialist career track. (We do not pay people for being able to make or support good statistics, we pay them only for their educational level and their management level). 3. Work consciously with motivation.
1. Lack of carrot and stick 2. Lack of motivation	1. Cannot reward people anything even for long, stable, good performance. 2. Cannot fire people, even if they don't work. 3. In fact, we punish people for good behaviour. Those most in demand get to do more work, have difficulty going on vacation, and when training offers occur, we often send people whom we won't miss when they are gone. 4. Both this and the fact that some people don't work much tend to demotivate even the good people.	

10 Appendix B. Other suggestions and recommendations

The most urgent, DISI specific recommendations were outlined in appendix A. The current appendix attempts to summarise the main suggestions from the main text but also adds a few that did not fit into the main text. As with most of the suggestions of the previous appendix, progress on most of the present suggestions are contingent on top management support and follow-up.

1. Have top management take a more direct interest/role in execution and follow-up. Too many initiatives seem to be left dangling somewhere in the organisation for no apparent reason. Some of these should be thrown out, of course, but some should not, and for these initiatives better follow-up procedures might get things moving.

2. Move main focus from output to capacity building[6]. Given good organisational structures and processes and a well-trained personnel, that is, given good individual and institutional capacity, one can move from fire fighting to execution of well-planned 'strategies'. Output will follow.

3. Move focus from short term to medium term: It takes time to train people, as does experimenting with and testing new solutions. More patience and longer term planning is advisable.

4. Make a long-term strategy that matches the expected (committed) financial and human resources

5. Be serious in applying the 3x3 rule. We are vulnerable if staffers leave. At the present time, for instance, only one person knows how to work with the Portal and the Dissemination Database. Also the infrastructure group is somewhat vulnerable.

6. Be sure to get competency groups above 'critical mass' and have them work as autonomously as possible.

7. The best staff is most likely to be the ones attracting offers from the outside so providing them good working conditions and interesting tasks is important.

8. Stop blaming the DPINE for not being able to deliver on their tasks, show them how to do it. The director for coordination has firm convictions as to how this can be achieved and might want to try a stint as delegada at a provincial office in order to the prove the case.

9. Give the delegadoes their own (dedicated) director to have their voice heard inside INE.

---

[6] This also goes for the current sponsors of INE. Focus in principle is on capacity building, but the big budget items are for 'output' - primarily for surveys.

10. Give DISI its own (dedicated) director, or move it to DCI

11. Introduce more systematical approaches to quality management (QA).

12. Introduce more systematical approaches to people management. Have people work (only) on approved tasks (approved by their most proximate superior). Maintain lists of approved tasks and register time usage accordingly. Use this data to measure performance and thus price of products, tasks, and services.

13. Have DISI, including the head of department and the adjoined director, have a meeting every day at 7.45-8 for making sure that everybody knows what they (and everybody else) will be doing that day, i.e., to summarise activities and get feedback.

14. Have all groups put a summary of the week on the intranet for sharing the state-of-execution.

15. Keep pushing for stricter (discipline in) adherence to IT policy, at all levels of the organisation.

16. Keep simplifying (homogenising) IT installations - especially at the DPINE.

17. Get more serious about physically securing the INE servers (including those on 9th floor that host survey micro data). We were lucky that were not taken completely out of business when the when the water pipes in the server room broke. (Please ascribe it to someone working on a Saturday). In the future it would be better not to rely on our good luck.

18. Get hardware for creating test systems

19. Allocate time for experimentation.

20. Create the Data Warehouse prototype. Make (keep) it simple. Use human readable formats. And gradually push the ownership to the statistics departments.

21. Make an in-house analysis unit in order to have its needs drive improvements in quality and accessibility of data.

22. Sell consultancy services, not (base) data or electronic versions of publications.

11 Appendix C. Problems in addressing IT Infrastructure problems

When arriving at INE, I was presented with three major problem areas; the portal, the DPINE networks/email, and the INE servers. In one of these areas, that of the DPINE, problems remain. In another, the INE Servers, we are only migrating as I am writing this. This is contrary to my expectations that all technical problems would be solved in less than a year (with me busily trying to figure out what to do with the rest of my term at INE). But then, almost no problems in this world only have a technical side.

DPINE

The most serious technical problem probably is the state of the DPINE IT infrastructure. In fact, the DPINE IT environments shouldn't be working at all. A network of approximately 10 computers ranging from new to obsolete, running many different operating systems (all Windows), with too many printers, all connected with 'hobby' network equipment, and linked to the internet by a modem. Not to mention the support technicians being up to 2000km away with next to no resources allocated for maintenance. One should be positively surprised that the DPINE IT is at least partly functional.

But - especially if one is serious about decentralisation - the DPINE IT infrastructure is certainly a serious problem. And is correctly considered so. In 2005, we made a large effort at getting the resources needed to improve the situation, but were unable to drum up support for the project. The two missions (Inhambane, Gaza) used for getting our act together - creating a model environment and thinking seriously about user training - were financed by the LTA local travel account, not by INE. And for the next three DPINE slated for an overhaul we succeeded neither in getting support for travel nor for buying the five computers needed.

Still, in the beginning of 2006, there was harsh critique of the fact that nothing was happening. So the DPINE was made the number one priority and money was allocated in the DISI budget for 2006 for updating all DPINE. However, there seems to be no direct link between the DISI budget and what money is made available for, and, to my knowledge, no purchasing procedure was initiated. It became still more interesting when the Italian project subsequently supplied us with a number of computers (and some DPINE found other sources for equipment) - in total, enough computers for overhauling all DPINEs outside of Maputo and Beira. One should then expect that funding for the needed networking equipment (less than USD 5000, for all DPINE) and for travel to the DPINE would be coming - after all this is peanuts compared to the census, and one certainly would like to have the infrastructure in place before then. However, after half a year of arguing about the true state of affairs, with DISI saying that things mostly work while everybody else said they did not, things reversed - with DISI wanting to go but being told that only Nampula had a problem serious enough to warrant travelling. And so, nothing have changed but the problem has gone away - at least for now.

It is important to note that - even if one does not decentralise, and even if the tasks of the DPINE are not considered all that important, and their productivity low - this does not entail the possibility of having a substandard infrastructure. As soon as one has an IT infrastructure, it needs to be maintained. And the worse its state, the more maintenance is needed. Computers need to be maintained whether used productively or not.

INE Servers
At the time of writing, the migration is at long last occurring. Here, it should be noted that it has taken two years - of hard work - to secure the resources for the migration. In spite of the fact that the migration follows logically from the choice of platform, detailed in the approved IT policy and strategy.

The migration has been well prepared, there was amble time for that. But it would have been relevant to have a thorough discussion, across all of top management, of wishes and needs regarding the future IT infrastructure. This should have yielded a
(small or big) revision of the policy and strategy. With the discussion of 'mini computers' (today, normally replaced by server clusters) cropping up, I am not sure where INE will be going, and if we took the right path. But if we are to choose a qualitatively different path, then the right day for starting the preparations is day after the migration has taken place - to ensure that there is time for experimenting, testing, and training. (Refer to the document on IT architectures for four possible main paths for the future).

Portal

Tellingly, the only problem considered resolved is that of the Portal. Subject to many concerns and dead ends what eventually resolved the problems was a reduction of image sizes (putting content over form) and increasing the bandwidth for the portal server (i.e., supplying the money needed for success). Technically trivial, but necessary.

12 Appendix D. 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 National de Estatística (INE) Maputo with possible travel up-country to Provincial offices