

BMB - LDK
Programme Monitoring Consultants to the
Office of the Minister of State for Administrative Reform
Beirut - Lebanon

Memorandum

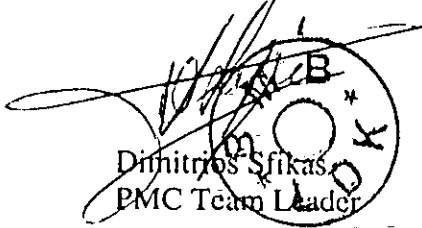
To: Dr Raymond Khoury, TCU Director, OMSAR
Mr. Atef Merhi, IDU Acting Director, OMSAR
Mrs Roula Kabbani, EU Project Coordinator, OMSAR

From: Dimitrios Sfikas, PMC Team Leader

Date: 25 February 2002

Subject: Mission Report to design the database of the CAS

Please find attached for your review and approval the Mission Report to design the database of the Central Administration of Statistics of the Short term Expert for Database design Mr. Arij Dekker.


Dimitrios Sfikas
PMC Team Leader

الجمهورية اللبنانية
مكتب وزير الدولة لشؤون التنمية الإدارية
مركز مشاريع ودراسات القطاع العام

Republic of Lebanon
Office of the Minister of State for Administrative Reform
Center for Public Sector Projects and Studies
(C.P.S.P.S.)

ARLA Programme

(LBNB7-4100/IB/97/0687)

Document for discussion, and not binding the OMSAR/PMC or the EC

**Assistance to the Central Administration of Statistics
(CAS)**

**Report of a mission to design
the database of the CAS**
January 26 – February 24, 2002

Programme Monitoring Consultancy

Arij Dekker

Short Term Expert for Database design

OMSAR, Beirut, February 2002

Summary:

The technical specifications of the existing Business Register at CAS were examined. The present design meets reasonable requirements and has served the CAS well. There is no urgent need to apply modifications. Nevertheless, this report formulates a simpler model that might be used in future, following a 2002 Buildings and Establishments Census. The linkage to geographical information systems – maps for which are already partially available - has been considered. Other front-ends to the Oracle database were discussed, including MS Access and Oracle Forms. In cooperation with CAS data processing staff a prototype database application was developed, which requires additional work.

The CAS situation in terms of computer equipment, software, and associated training, was assessed. Proposals for prompt action for improvement were communicated to the PMC and reviewed with the responsible specialist, prior to implementation. There is urgency here, since the planned Census requires these resources.

Terms of reference

Proposed specific objectives of the mission

The specific objective of the proposed assignment is to review the design of the existing Oracle database at the CAS and enhance the design to improve its functions as sampling frame for household and economic surveys respectively, and as a basis for structural statistics on buildings, dwellings and economic establishments.

Main activities to be undertaken

In cooperation with the CAS staff the Database Expert's main activities will be to design the Oracle database of the CAS. The specific tasks of the expert will be to:

- a) Analyse the existing database in terms of its several functions in the statistical production system of the CAS. This will include identification of the current data model of the database (tables, indexes, constraints, views, relations, ...)
- b) Design the suggested database in terms of the required data model as on both logical and physical level, the migration roles between the present and required databases, the network and the platform for the system in terms of hardware and software and storage and processing constraints in terms of volume of data and transactions and number of users and expected active connections.
- c) In co-operation with the geographical database expert, establish the connection between the geographical database and the Oracle database.

Expected outputs

- A global report including the tasks fulfilled, particularly,
- The complete data model of the existing database;
 - The complete design of the required database;
 - The migration rules;
 - The geographical database connection rules.

Tasks undertaken which were not identified in the Terms of Reference

As is not uncommon in this type of assignment, there were several other tasks to be undertaken. This concerned primarily an inventory of the requirements of the CAS in terms of computer hardware, software, and associated training. An equipment tender had to be issued urgently, and training needs were to be formulated. The adviser took care of this in cooperation with Messrs. Abdallah and Catinis. There was also the need to prepare a work plan and a budget for the soon to be undertaken 2002 Census of Buildings and Establishments. Finally some hands-on training was given in front-ends to the Oracle database system, in particular MS Access via ODBC connections.

State of affairs

The current situation at CAS has been described in some detail in various reports prepared by Prof. Sten Johansson. Suffice to mention here that CAS is a young institute that came into being only in the early nineties, after the civil war in Lebanon had come to an end.

Presently the country still lags in statistical development as compared to nations of similar size and economic conditions. The CAS has a permanent staff of less than 100, counting professionals as well as general service personnel. With many of them rather new to their jobs, statistical production is still small and focussed on priority needs such as periodic publication of the consumer price index and responding to various information requests, many from international agencies. CAS also acts as a clearing-house for sector statistics obtained from other government organisations and the central bank. Due to the scarcity of indicators on the national production of goods and services, as well as about household incomes and consumption, preparation of national accounts of an acceptable level of reliability is still a problem. This is now a priority development area, supported by the current project. For the year 2002 an in-depth economic survey of establishments is planned, as well as the Census of Buildings and Establishments already referred to. Early 2003 will see a household sample survey. The two surveys will cover the production and consumption sides of national accounts, and are expected to provide a good basis for estimates of the GDP and other principal economic indicators.

There are now some 20 PC's of reasonable performance (better than Pentium/200 MHz) installed at the CAS. Furthermore, the office has a rather outdated UNIX Server (Bull DPX/20-300, PowerPC processor, AIX operating system), and an office-wide UPS system. At the time of arrival of the consultant the UPS system had been out of action for some time, due to defective batteries. As a consequence, the server also was not operated, for fear of it being damaged by a jolt in the inconsistent Beirut public power supply. These problems were solved in the course of the mission, thanks to emergency support from the PMC.

The existing Establishments and Buildings database

The existing database results from a 1996 national census (which then excluded the occupied Southern territories). While some further data gathering has taken place, which has resulted in additional data sources, the 1996 database has not been systematically updated. This is due to the difficulty of merging old and new data files. In future this will become easier once, as intended, each building will have x,y co-ordinates among its attributes. These co-ordinates will be obtained from digital maps for urban areas that CAS now has been provided with, and by using handheld GPS (global positioning systems) units in the rural areas of Lebanon.

The E-R diagram of the existing database is shown in Appendix 3. Primary key fields are shown in capitals, foreign keys are at the end of 1-∞ (1-n) connector lines. Page Appendix 3.1 depicts the *geographical entities*, which are completely hierarchical in nature. The administrative units are Muhafaza-Caza-G(rand)Zone-Circ(onscription)Fon(cière)-P(etit)Zone. The Ilot, the lowest geographical level, is the CAS-designed statistical base unit. Ilots fall within a single Pzone. There presently are some 11,499 of these basic enumeration areas. Ilot boundaries will be redrawn for the new Census. Recorded x,y co-ordinates for the centres of the various administrative divisions allow in principle thematic maps to be generated from the existing database.

Appendix 3.2 shows primarily the Bldg (Building) and BldgUnit (Building Unit) entities, making up the national *building database*. BldgAddress and BldgAppl store additional attributes of the Bldg table, and may have been split off for reasons of retrieval efficiency. The design seems fine, although there may be some dependencies within rows that would violate the third normal form (e.g. in the table BldgAppl: TotalUnits would be HomeUnits +

NonHomeUnits + InstHomeUnits + UndefUnits). Storing this sum - that could otherwise be easily calculated - as a separate attribute may result in inconsistencies.

Page Appendix 3.3 shows the national *establishments database*. While there seem to be many tables, in fact most store attributes of InstMainInfo and have a 1-1 relation with it. Therefore this database too holds only two principal tables, those for Establishments (InstMainInfo) and Branches (InstBranches). Ancillary tables are the ISIC (International Standard Industrial Classification) codes and some of the geographic entities (Muhafaza, Pzone, Ilot).

The main office and branch offices of an organisation are establishment units in themselves. Therefore it seems not entirely logical to store them separately and with different attributes. The apparent reason is the method of field data collection in 1996, whereby enumerators asked from each establishment questions about the main office (if any) and the branches (if existing). The responses were recorded as such in the database. In most business registers main offices and branch offices are stored in a common establishments table, with pointers for each branch office referring to the applicable main office. Suitably defined indexes will facilitate retrieving branch offices depending from a particular main office.

Here again there are some dependencies between non-identifying attributes (PzoneSN – PzoneCode in the InstBranches table) that could have been avoided, since there is already a link to the Pzone table which can provide the PzoneCode for each PZoneSN.

The largest tables are Bldg (518,851 rows), BldgUnit (1,456,463) and InstMainInfo (198,446). Surprisingly, there are more main offices (InstMO: 6,001) than branches (InstBranches: 4,416). The number of BldgUnits is more than one would expect for Lebanon, although there are currently no reliable figures for the total population or number of households resident in the country.

Retrieval from the database is reasonably fast. CAS can obtain such information, as it requires, provided that the information exists in the database. Therefore the database is useful and constitutes an important asset to the CAS. However, the lack of updating makes the available information increasingly obsolete.

The new Establishments and Buildings database

The contents of the new database will depend on the available information, in particular the contents of the field questionnaire that will have to be completed by the enumerators. At the time of writing the design of this questionnaire had not been finished, but it looks as if it will be less comprehensive than the one used in 1996. Thus the database tables will be smaller in terms of number of attributes per entity. A tentative E-R model, shown in Appendix 4, occupies only a single page. The geographic entities remain the same (although Ilot boundaries will be redrawn). The other main entities: Buildings, BuildingUnits and Establishments, each are stored in a single table. The Establishments table contains internal pointers to link branches to their headquarters. One should note that this could not be the other way around (headquarter records pointing to branches), since that would introduce a repeating field, not normally allowed in relational database design. There are ways to solve this problem, but discussing that issue goes beyond the purpose of the present

report. Referential integrity of this internal link cannot be reinforced (establishing a necessary n-1 relationship between subsidiaries and their common head office) as long as not all head offices are represented in the database. The E-R diagram therefore shows a non-enforced relationship.

Just as in the 1996 database, the main tables have primary keys that are automatically generated whenever a new row (also called: tuple) is added to a table. This mechanism is known as AutoNumber in MS Access. Using indexed keys without intrinsic meaning prevents problems when tuple codes (such as building or establishment identification codes) would be subject to modification later on.

The principal constraints are those imposed by the foreign keys: building units and establishments must be linked to one and only one building, and all ISIC codes have to be existing. Furthermore, there are obvious range constraints in other codes and attribute values (legal status of establishments, number of rooms in building units, starting year of establishment, ...).

Due to the difficulty of reconciling newly recorded buildings and establishments with their earlier records (if existing), there can be no meaningful migration of the old data into the new database. Thus, the issue of identifying migration rules is largely void. This has for consequence that historic information will be unavailable at the individual tuple level. However, it remains possible to enter accumulated 1996 data in the new database. This could be, for example, total numbers of establishments, buildings and building units, and totals of various sub-sets, at the administrative levels used in 1996 and to be used again in the new Census. Such historic totals would be highly useful for analysis.

For the front-end to the Oracle database there are a number of possible choices: Oracle Forms, MS Access, Borland Delphi, programmed Visual Basic interface, ...). All would perform adequately if designed properly. The combination MS Access - Oracle has some advantages. The front-end cost is comparatively low, and development work can be done in a complete Access environment, for subsequent upgrade of the back-end to the more secure Oracle. During the mission some development for the new database was already undertaken, using these tools. A currently available utility that comes with newer versions of Oracle RDBMS allows automatic upgrading of Access databases (including their forms) to move the tables to Oracle and adapt Access forms to manipulate the now remote data.

Together with Messrs. Abdallah and Brown, the adviser reviewed the necessary communication link between Oracle 8I and ArcView 3.1, currently both available at the CAS. Using Oracle 8I as an ODBC (open database connectivity) data source for ArcView 3.1 was shown to be functional both on a stand-alone system and through the local area network. ArcView 8.1, and the latest versions of ArcGIS, can also employ data sources observing the OLE DB (on-line linking and embedding for databases) protocol. The combination ArcView 8.1 and Oracle 9I (both not yet installed at the CAS), does this well. —

Entities to be shown on maps obviously require co-ordinates, as points, line segments, or polygons. Adequate connection between the software systems can be then obtained via commercially available ODBC and/or OLE software links. It is intended that the new database will have co-ordinates x,y for each building. These would be obtained from now available detailed maps for urban areas, and enumerators using GPS in rural regions. The coordinates will allow individual buildings – and the establishments and dwellings

associated with them – to be shown on the maps.

A mechanism should be developed to regularly update the new database once it has come into being. If not, it will soon lose its effectiveness as a source of information and as a statistical frame for drawing samples for business and household surveys.

Other activities undertaken during the mission

In co-operation with Messrs. Ziad Abdallah and Lian Catinis the adviser drew up specifications for the immediate requirements of CAS in terms of computer hardware, software, and IT training (Appendix 2). These will be addressed via one or several equipment tenders and a PMC framework training agreement. It is hoped that the first procurement will be completed within four months. The items are very much needed for the purpose of undertaking the Buildings and Establishments Census scheduled for the present year.

A time plan for the 2002 Buildings and Establishments Census was also prepared, using MS Project software. It details the time and manpower minimally required for each of the various census activities. The Census is supposed to be completed early enough to use the results for drawing a sample of dwellings for a national household survey to be conducted in early 2003. This imposes a tight time frame, which will require a great effort from all concerned.

Conclusions and recommendations

The CAS has embarked on an ambitious programme for 2002 that requires all potential resources to be mobilised. This mission looked into the IT aspects, and database design in particular. While much remains to be accomplished, it appears that the fundamental pre-requisites to make the efforts a success are fulfilled.

Recommendations are as follows:

- The tendering process to obtain hardware, software and training should be closely monitored, to make sure that no undue delays occur;
- The questionnaire for the 2002 Buildings and Establishments Census should be finalised as soon as possible, to allow IT development to proceed;
- Upon completion of the questionnaire design, the new database model can be completed and a front-end permitting efficient data entry may then also be finalised and put to the test.

Acknowledgement

The consultant would like to express appreciation for the kind hospitality and full co-operation received from CAS as well as the ARLA PMC.

Annex 1. Contacts**Central Administration of Statistics**

Ms. Maral Tutelian	Director General
Ms. Najwa Yacoub	Statistician
Nader Keyrouz	Statistician
Ziad Abdallah	Data processing specialist
Daher Doumit	Statistician
Ms. Elissar Nasser	Analyst
Ms. Siba Haidar	Analyst
and other members of the CAS staff	

Technical assistance team, ARCADIS BMB

Sten Johansson	Team Leader
Hilmi Dabbagh	Economic statistics
Robert Brown	Geographic information systems
Ms. Zoha Khalil	PhD statistics (local expert)

Project management (Beirut)

Dimitrios Sfikas	PMC Team Leader
Marten van Golverdinge Schut	Project Management
Ms. Rola Alameiddine	OMSAR counterpart
Lian Catinis	IT Expert
and other members of the PMC staff	

Project management (The Netherlands)

Pieter Feenstra

Annex 2. Equipment and Training requirements

11/2/02 CAS

Requirement IT equipment (tentative)

Oracle server

Min. specs: Pentium III Xeon 550 MHz, 1 GB RAM, 100 Gb disks (SCSI, RAID 5), W2000 Adv. Server

Oracle is the principal software tool at CAS. A dedicated server is preferred, both for availability and security purposes

Application server (also for GIS)

Min.specs: Pentium III 600 MHz, 512 Mb RAM, 60 Gb disks, W200 Adv. Server

This server will provide application software other than Oracle (ArcGIS, SPSS, specialised statistical systems)

Intranet server

Min.specs: Pentium III 600 MHz, 512 Mb RAM, 60 Gb disks, W2000 Adv. Server

This server is to host Intranet services that will allow CAS staff to work out of office

It is physically separate, to cut off any access of hackers to sensitive CAS databases

Server for file storage and shared peripherals

Min.specs: Pentium III 600 MHz, 512 Mb RAM, 60 Gb disks, W2000 Adv. Server

This server will be host to shared peripherals: volume printer, fast scanner, ...

Quantity	Unit price €	Proc. phase	Cost estimate	
			Phase 1	Phase 2
1	10,000	1	10,000	
1	4,000	1	4,000	
1	4,000	2		4,000
1	4,000	2		4,000

الجمهورية اللبنانية
مكتب وزير الشؤون الإدارية
مركز مشاريع وتدرجات القطاع العام

Ethernet switch Min. specs: 12 ports, stackable The CAS network currently operates with outdated hubs, which generates unnecessarily traffic on many connections	8	250	1	2,000
PC type I Min. specs: Pentium 4 or equivalent, 2.2 GHz, 1 Gb RAM, 60 Gb hard disks, Ethernet, sound, CD read/write, 17" monitor, W2000 Top-range PC's for the IT staff	6	2,200	1	13,200
PC type II Min. specs: Pentium 4 or eq., 1.5 GHz, 256 Mb RAM, 40 Gb hard disks, Ethernet, sound, CD read/write, 17" monitor, W2000 Advanced PC's for professional staff (existing equipment goes to data entry, etc.)	16	1,400	1	22,400
PC type II Min. specs: Pentium 4 or eq., 1.5 GHz, 256 Mb RAM, 40 Gb hard disks, Ethernet, sound, CD read/write, 17" monitor, W2000 Second batch of advanced PC's	8	1,400	2	11,200
PC, Notebook Min. specs: Mobile Pentium III or eq., 1.0 GHz, 256 Mb RAM, 20 Gb hard disk, Ethernet, CD, 15" screen, W2000 For censuses and surveys CAS staff work throughout Lebanon. They need notebooks to allow data processing while out of office.	4	2,500	1	10,000
PC, Notebook Min. specs: Mobile Pentium III or eq., 1.0 GHz, 256 Mb RAM, 20 Gb hard disk, Ethernet, CD, 15" screen, W2000 For censuses and surveys CAS staff work throughout Lebanon: they need notebooks to allow mobile data processing	2	2,500	2	5,000
Printer, central Min. specs: 16 ppm color laser This printer is mainly intended to produce small series of official reports and press releases	1	5,000	1	5,000

Printer, individual Min. specs: 4 ppm color jet Printers to be shared by several individual users	5	500	1	2,500
Scanner Specs: High-speed duplex scanner with ADF (Automatic Document Feed), A3/A4, 200 dpi, 30 ppm (duplex) To establish an electronic archive (imaging questionnaires, etc.). Currently much valuable office space at CAS is occupied by document registries. As the staff is expanding, this space is urgently required	2	4,000	2	8,000
UPS Min. specs: 1.5 KVA, 6 outlets, USB control port To decrease CAS' dependency on an increasingly unreliable central UPS (Note: Alternatively a central 20 KVA unit and two portable 1.5 KVA units may be provided)	10	700	1	7,000
GPS receiver, handheld Min. specs: waterproof, accuracy 15 meters, shows longitude/latitude To be used by field workers, e.g. in field work for the 2002 establishments and buildings census CAS is building a GIS in which each building in Lebanon will have its geographic coordinates attached	200	150	1	30,000
Beamer Min. specs: resolution 1024x 768 (XGA), 1000 lumens For in-house training and presentations to data users	2	3,500	1	7,000
Equipment costs				113,100
				32,200

Note: all equipment to be warranted for 3 years

Requirement software (tentative)

Upgrade Oracle DBMS from version 7 to latest available in Lebanon

Developers: 6

Users: 20

2

p.m.

2

p.m.

Upgrade Oracle Developer to latest version

Developers: 6

2

p.m.

ArcInfo primary floating license

1

4,500

ArcEditor 1st License

1

9,000

Upgrade ArcView 3 to ArcView 8.1 Concurrent

1

4,800

Software costs

18,300

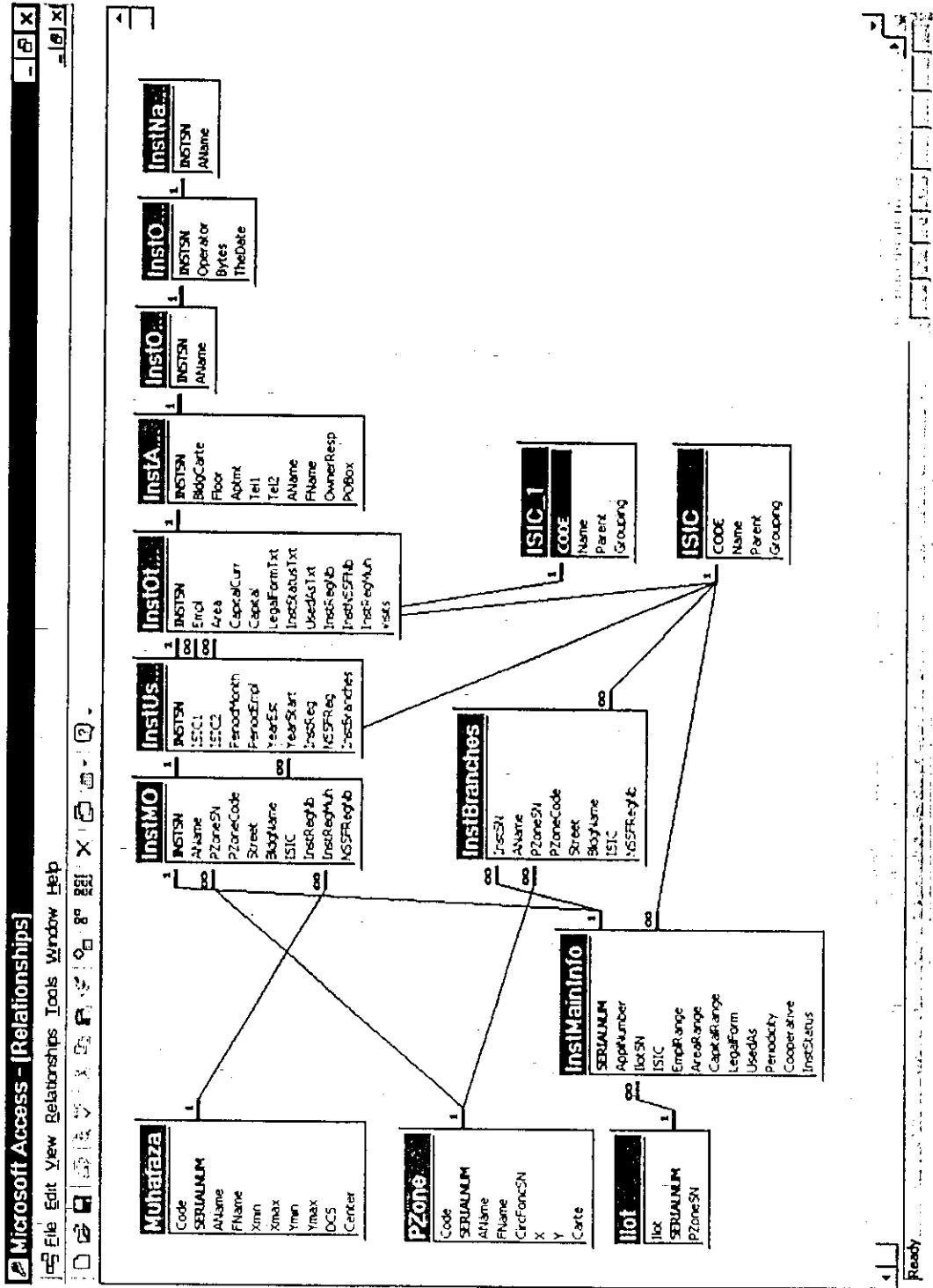
0

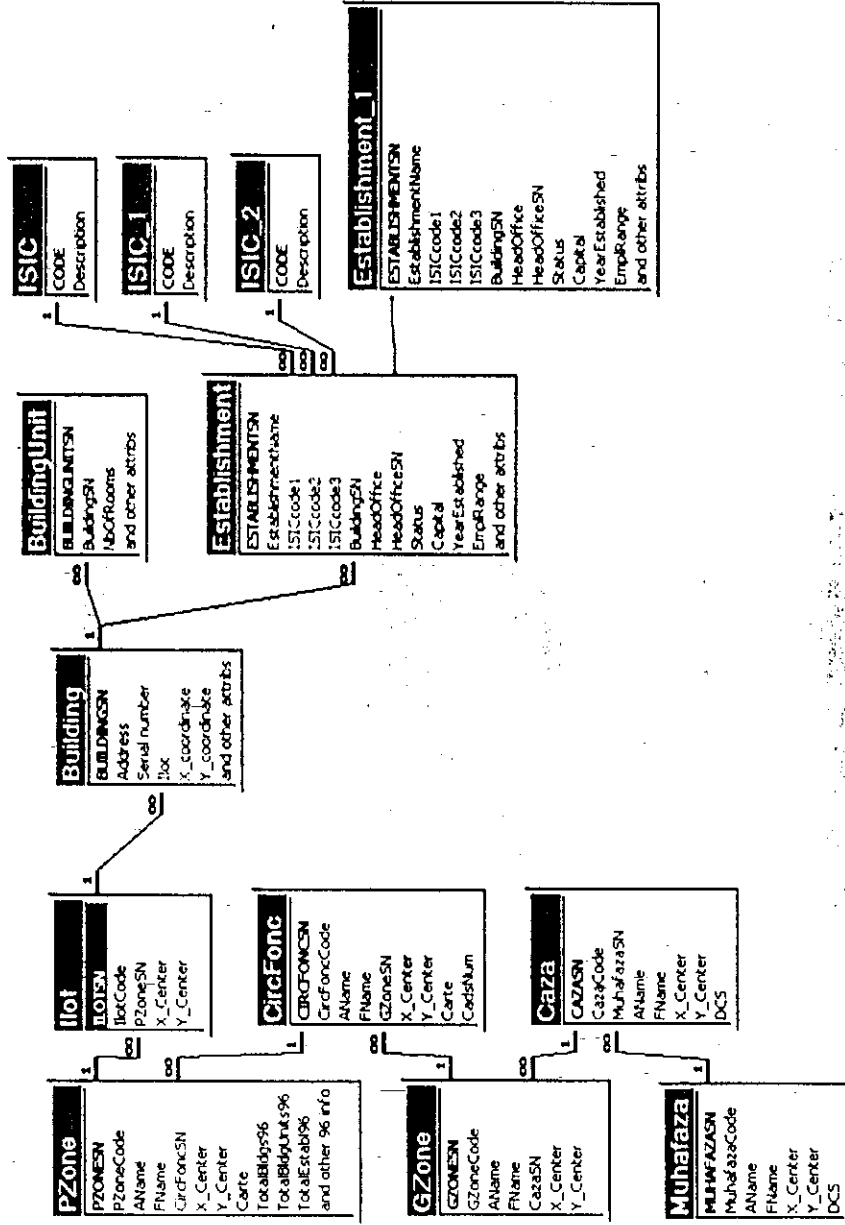
Total

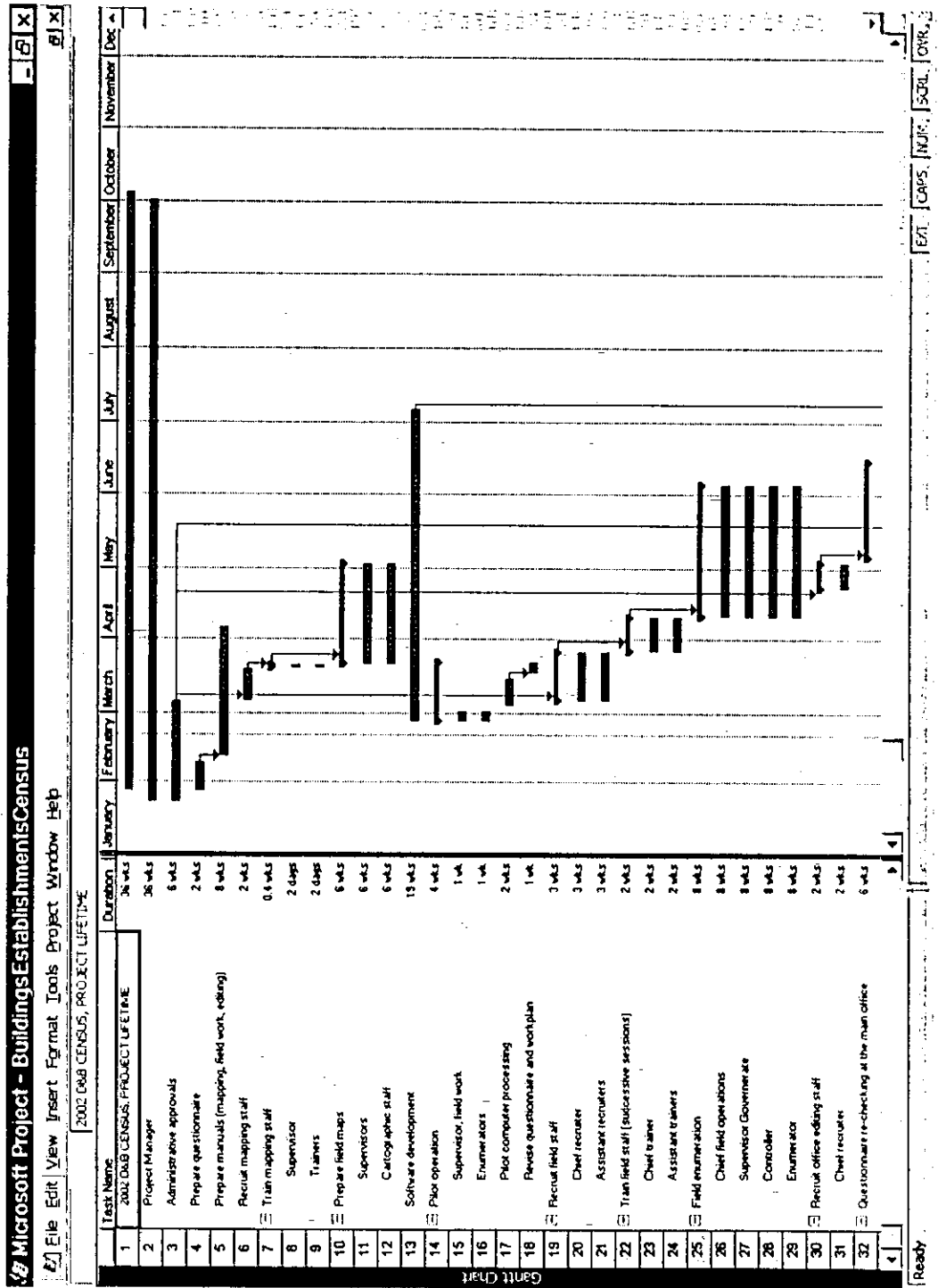
131,400	32,200
---------	--------

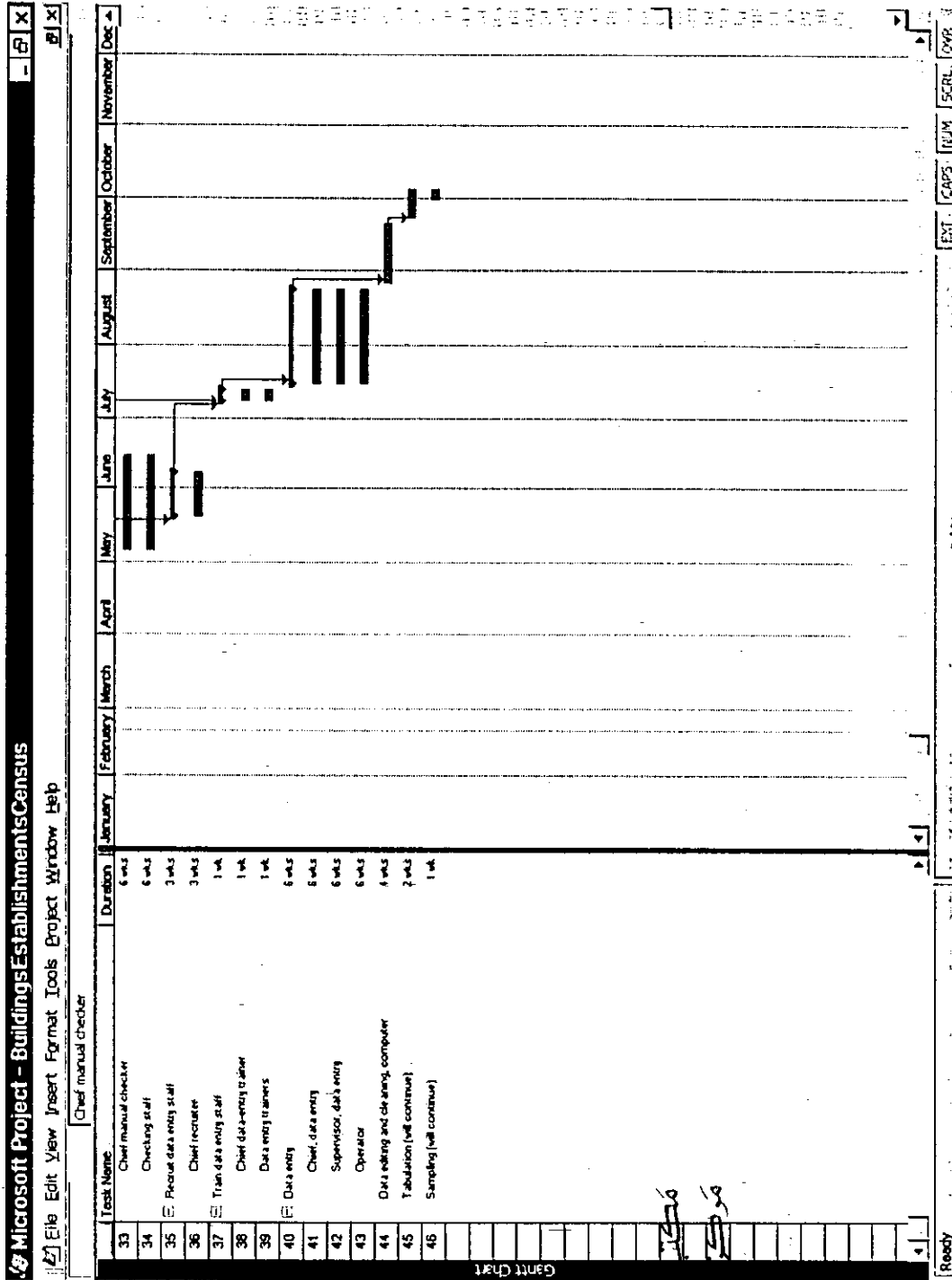
Requirement IT training (tentative)

	Attendees	Units
<i>Oracle applications</i>		
Introduction to Oracle, PL/SQL program units, building applications in Forms and Reports	6	21
<i>Oracle server administration</i>		
Architecture & Administration, Backup & recovery, performance tuning, network administration	3	19
<i>Oracle designer</i>		
Data modeling & relational systems, systems modeling, forms and server generation	3	9
Microsoft MCSE (Windows 2000/network management)	6	
Visual Basic	6	
ArcGIS module 1	6	
ArcGIS module 2	4	
Statistical software (SPSS)	12	









مركز الدراسات والبحوث الإدارية
 جمهورية اللبنانية
 مركز