EXTERNAL AND INTERNAL EXIGENCIES ON DATA QUALITY IN DMS

Lujan RUIZ DIAZ
Edenor S.A. - Argentina
Lruiz@edenor.com

INTRODUCTION

This Paper deals with the complexity of the information systems and data quality, integrated in a DMS (Distribution Management System) in the electric utilities that operate under the de-regulated market normative. The exigencies of the Governmental Control Authorities to audit the tracking of the supply quality (mainly interruptions) are also detailed. The problems derived from the high quality level of the data required by these external exigencies without a real necessity, because not all technical data needs to be 100 % error free, becoming a heavy burden on the utility budget.

FRAMEWORK DESCRIPTION

First, it is necessary a brief description of the utility under analysis to give an idea of the general frame, and to understands the mainly aspects to deal with. The distribution utility Edenor S.A., covers supply in the northern Buenos Aires City and surroundings, in Argentina. It covers supply over a surface about 4,637 km², (almost 8,000,000 inhabitants), with approximately 2,250,000 customers, an annual energy surrendered to the network of nearly 14,865 GWH and a maximum demand of 2,800 MW (year 2002). It possesses 63 Substations HV/HV and HV/MV, and approximately 13,500 MV/LV transformation centres, 1,200 km of HV network, 7,780 km of MV network, and 25,000 km of LV network.

Their influence area goes from very high load density into Buenos Aires city, with buildings of 40 or more floors, up to the countryside areas with MV feeders 50 km long or more. The Paraná River Delta Zone has a very low customer's density, with forested islands where the only possible access is by fluvial sailing.

Because of that, it is evident, that Edenor faces a wide range of troubles to homogenise the technical and commercial data needed to fulfil the internal and external requirements of the information systems.

GENERAL REQUIREMENTS ON DMS DATA

At present time, the Governmental Control Authority (ENRE) exigencies are to tracking and audit the supply’s continuity in each one of the utility customers. With that, the LV level is also included. In summary, all the interruptions that affect to any distribution utility customer (from the biggest company to the smallest residential) should be tracking and know [1].

It is mandatory to calculate the Energy Not Supplied (ENS) out of quality limits, because the algorithm for monetary penalties (that are refunded to the affected customer), take in account this value. The calculation algorithm of the ENS is variable in function of the customer's rate, annual energy, the curve of standard demand of the same one (fixed by the Concession Contract), and the time when interruption occurs. The previous interruptions are also a necessary input to the calculation algorithm. Because of these requirements, a record of each customer's interruptions should be known, it causes and schedules of each one of them.

The problem is not the complexity of the algorithm (it is only mainframe calculation power), once you agreed with the Control Authority some subtle interpretations, but the necessary data and his quality level.

The control organism information’s demands are very detailed. This makes that in a six months control period, should be given around 100 000 000 data, many of them very complexes (files with more than 12 000 000 records). In many cases, the joint of other fields from different databases composes these data fields. By example, the customer electrical path, that includes codes of: HV/MV Substation, HV/MV transformer, MV substation bus bar, MV feeder, MV/LV centre, MV centre bus bar, MV/LV transformer, LV feeder, and customer LV delivery point. The same kind of data is required in each interruption to inform the operated switchgear electrical path.

Of course, the most part of these data are dynamical to the minute, due to the network operation, e.g. the momentary configuration changes to operate the network and restore the service, by means of alternative feeders. This poses additional problems to maintain the data quality.

At Edenor, the whole of this technical and commercial information is stored under the name BDE (Exploitation Data Base), which is on Oracle platform. However, BDE only receives the data from others information systems (SCADA, commercial, customer’s claims, technical subsystems, legal documentation to support “force majeure” cases to avoid penalties, etc.). The systems diversity poses one of the main problems on data quality.

The data are a company asset, the same as the installations, and nobody discusses this fact. The data are the raw material to be process by the systems and the final product (decisions on network operation, maintenance, planning, investments, etc.) reflects the initial data quality level.

To maintain a very high level of quality and availability is mandatory in certain systems. A SCADA required nearly a
few millisecond precision and instant availability to avoid perhaps a general blackout. The main question remains and it is: all technical data needs to be 100 % error free?

DATA DEFINITIONS

It is not necessary to repeat here all the theory and real aspects about the data quality, which has been previously enunciated in many forums, proceedings, articles, etc. The reference [2] contains a fine resume and subject description.

In a general sense, and not only in distribution utilities, all people agree enthusiastically with the principles, but to synchronise all the gear to work appropriately is another matter. Finally some type of “Not in my backyard, please. It is your business” attitude prevails, due to the not understanding of the other people and applications necessities, and to the lack of a corporate focus and commitment on the subject [3].

Three common examples on problems on technical and commercial data in a DMS are given:

Different Definitions On The Systems

There are many common data with a particular definition in each information system, but the same data name does not imply the same meaning. A simple example: “customer”:

For a commercial agent, “customer” is an individual or company, which have a commercial relationship with the utility. Perhaps this individual has moved to another address, but the economical issues are not been cleared yet. With the commercial approach, the account number exists and the “customer” concept is valid, possibly with a special mark at the database.

For the technical agent (e.g. the one involved with the service quality and interruptions tracking), this “customer” does not exist anymore, because for him, a customer is an individual with a meter and connected to the electrical network.

Both definitions are “correct”, within the boundaries of each system (commercial and technical). However, the problems surge when the systems interchange information, in this case, typically, the date that ends the relationship between customer and utility. The technical system needs the date when the meter is disconnected, and the commercial one, the date when the debts are settled.

If the signal from the commercial system to the technical one, to de-activate a customer, is this date, then, there are problems to count and tracking the interruptions. This implies errors with the algorithm to calculate the eventual penalties to be refunded to the customer.

Mismatches Formats Or Structures

In other cases, there are no doubts about the meaning of the data fields in different systems (e.g., the data field that identify a HV/MV Substation). However, perhaps in the SCADA subsystem, the data format is defined as a numeric code, and in the technical database as an alphanumeric field, that differs from the SCADA for the same object (the particular Substation).

An equivalence data interface is needed to link both systems, but the basic problem remains. Generally, it is not easy to change the basic definition of the data field in a system like SCADA.

The Data Warehouse concept and the ETL (Extraction, Transformation, and Load) tools may help to solve the mismatch.

The Reality Of Data Entry Level

Very often, the data definitions are made without a clear idea of the field problems at the data acquisition and entry level, because the system has restrictions and it is sell “as it is”. Any further attempts to fix the trouble usually get the situation become worse. A clear example is the customer’s address data fields. These fields are typically defined as:

Street – Door Number – Floor - Apartment

This is fine for the downtown customers, but what happens when the customer is in a rural area? Perhaps the street name exists but the number on the door does not.

A solution is to enter a dummy door number, but the risk to duplicate an address exists, except if a clear internal procedure is established, as at Edenor. Without such procedure, the people from “the front line” usually ask to the application supporter for a special character (e.g. a currency sign “$”), in order to force the data entry. Also to avoid the consistency control that demands a numeric value on the field “door number”.

Now the risk to obtain bad information at the data entry level worsens. If a table of possible streets names and numbers to control the entry is integrated to the system, people in the field can be tempted to force any number, even in the cases when the correct and real one exists. Certainly, the ideal solution does not exist. At Edenor, in the Paraná River Delta Zone, there are some actual addresses like “Capitán River, corner with Pacifico Creek, and 150 meters to the left from the black pier”, and it is necessary to adapt these
real data. The DMS systems evolved to the integrated GIS, and a solution based in GPS algorithms would be possible in these cases.

**UTILITY INTERNAL EXIGENCIES ON DMS**

In nearly every distribution utility each system or subsystem has been put in operation with different scopes, diverse dates, and various technologies.

**Edenor Situation**

At Edenor (owned by EDF international), some technical and commercial procedures run today under ISO 9000 quality certifications. The corporate and management commitment with the quality is very high, with the compromise to extend the ISO certifications to all the technical, logistic, and commercial procedures. Nowadays, the new SIGT (Technical Management Integrated System) is in the early stage of definition, in order to integrate and improve the existent systems and added new subsystems and technologies.

The need to integrated systems is unquestionable, and there are many ETL tools and middleware (software to support interactions) in the market, but the first step is to develop a Common Information Model (CIM).

**General Aspects on the Subject**

Fortunately, a great amount of work has been made with the approach to a CIM, under the auspices of the International Electrotechnical Commission (IEC), and some references are given at the end of this paper.

One great advantage of the IEC recommendations to facilitate the systems integration is that they are “neutral”, and can be adopted as reference by the company, with a minimum resistance from people and internal organisation.

The technologies evolve very quickly, and there are some application and projects [4], that are taking full advantage of the capabilities of the state-of-art.

Nevertheless, it is mandatory not be blinded by these capabilities. All systems need to be feed with data, and the cost to obtain the data in the field, and to maintain it, can be very high. The project on reference [4] is an interesting approach to a special solution needed by the utility, but the estimated cost to obtain the data is over the 10 US$/customer. The new dazzling interface technology (three-dimensional perspectives, colour layers, etc.) can be very attractive at the demonstration stage. However, it cannot be forgotten that the data availability and quality is the first and unavoidable step.

**Remarks On Data Acquisition and Maintenance**

The distribution utility internal exigencies on DMS must be harmonised, and a corporate focus and commitment on the subject must be launched immediately, even if the integrated project is in an early stage. It is possible also to initiate a Data Quality program, aimed to the Data Warehouse concept as a solution to integrate the various and diverse systems.

A cleansing method to the existent data must be put in operation. That means modify incorrect values, conform some existents to the company standard, or “fill the holes” with clear rules. In addition, assuring the quality at the acquisition and data entry level (prevention is the least costly solution) is a must. Another good procedure is to maintain only the necessary standard of quality (100 % without doubt in certain cases, but the absolute defect-free is not the general rule), along the full life of each data.

Before the decision to acquire a data, and with the assumption that it is well defined, some questions must be formulated and answered with intellectual honesty:

a) Is the data obligatory or optional for the systems?

b) Is it possible to acquire and load the data at reasonable cost?

c) Can be a validation routine implemented at data entry level and afterwards in relationship with other data from the systems?

d) Can be the data maintained easily? Which is the cost?

e) Which accuracy level is needed?

f) The sole data availability does not justify their inclusion on the database. A serious ROI (Return Of Investment) study must be carried out, because the initial data load and maintenance task can easily count for half the cost of an information system [3].

**Final Comments**

It is unreasonable to dump existing data (such a valuable asset, even with some errors in it) in order to start from scratch the implementation of a large, stand-alone, integrated system to fulfil all technical, commercial, legal, and economic needs. In the other hand, it is doubtful that such kind of system exists as a market solution. The existence of separate information systems will go on. The solution is to integrate them with clear rules in mind. This assertion does not prevent the existence of a main and reliable information and management system, which interacts with others subsystems.
The requirements for data should simply fulfill those of the systems that use it. If careful data definitions and a realistic quality level, in correspondence to people and applications' needs, are achieved, half of the battle is won.

EXTERNAL EXIGENCIES ON DMS

The distribution utilities like Edenor, operating under quality of service regulations (technical and commercial) and are controlled and audited by Governmental Authorities, faced also other troubles on the DMS data: the external exigencies.

In these cases, the DMS internal design becomes useless if it does not comply with the requirements of the external Authority, because it is mandatory to prepare and process the data, according with their exigencies.

In the Argentine case, the penalties that concern the data quality or process result can be very high. (So high, that it is possible that the amount surpasses the bonuses refunded to the customers, even if the non-fulfilment is insignificant). Some related constraints are briefly enumerated below:

It is mandatory to take in account all the external exigencies on the DMS design, even if they are not useful for the utility. These exigencies can be very costly and restrictive to the data model internal development.

Sometimes, it is necessary to develop and add to a market DMS, very specific and complicated interfaces to deal with the external exigencies.

The external data demands are generally aimed to audit some technical subjects under a pure legal focus. This implies unavoidable crashes between both approaches.

In Argentine, there are many Control Authorities (Federal and Provincial), nearly one for each utility, and theirs control models can differ. With these constraints, it is very difficult to the IT providers to develop a common DMS with all the needed functions. On the order hand, when a Control Authority has jurisdiction over more than only one distribution utility, the exigency is that all the companies must presented their data under the same model (Called Data Interchange Model), even if the utilities technical characteristics and systems differs.

Often the Control Authorities, from dusk to dawn, change theirs requirements, given a very short time to fulfill the exigencies, under a penalty menace. These changes in the external requirement can affects the DMS model very deeply. In consequence, all the emergency solutions (due to the time limit constraint) finally ended with a higher risk of poor data quality or wrong results from the data process.

Usually, the audit results have a significant delay (many months, even one year or more). If some minor systematic deviations are founded, the utility had no chance to correct them, until to know the result, and risk to pay fines for all the elapse months of data delivered with the systematic deviation.

The general rule is that when the legal scope prevails over the technical one, the external exigencies are data 100 % error-free. Then, high fines can be applied, because a clear criteria to evaluate the eventual deviations founded, does not exists.

Actually, the cited evaluation criteria must be defined taking in account the eventual data deviations on technical issues, or their influence on the bonuses to be refunded to the customer. Nevertheless, it is established that the penalties derived from data quality are “discretionary”, and evaluated and fixed by the Control Authority. The next two examples on data quality and data process are demonstratives of these facts:

**Time Discrepancies**

The discrepancy limits in interruption declared length has been fixed to one minute. This will be fine, and must be respected for the interruptions at the level of a MV feeder on a HV/MV Substation, and where a SCADA is operating.

However, when a crew is working in an isolated countryside area, and manoeuvring a switch or a fuse on the top of a pole, the situation is very different. Assume that they are restoring the service to four or five MV/LV transformers, with not more than 10 or 15 customers each.

If the crew has left the truck 200 of 300 meters from the pole due to access difficulties, it is necessary to return at this position. Then call by radio to the MV Control Centre and inform that the service has been restored. Of course, the Control Centre, ask what time it is. What response is obtained? Undoubtedly, an approximate data, from one of the men’s wristwatch, mixed with a personal consideration about the passed time, from the actual restoration to the radio call.

A difference of more than one minute with the reality is more likely, than to obtain the exact duration. In addition, if it is a typical interruption in a rural area, easily with three or four hours of duration, to inform 287 or 283, instead of 285 minutes, does not change the facts. Neither the eventual refund to the customers is affected (perhaps in a centime fraction, in more or minus, to be precise at the thousand).

The detailed case principles can be applied to nearly all the manual operations to restore partially the service on the network, where a SCADA does not exists, and time must be informed via a radio call to the MV Control Centre.

It is an example, where under a technical scope, the absolute data accuracy is not necessary, and to obtain the exact time, can be very costly. However, if the Control Authority had installed an audit event’s recorder, the distribution utility risks a high monetary penalty, even if the interruption is admitted by the fixed quality levels and not fined itself.
The aim of the quality of service control system must be to improve the quality of supply perceived by the customer (via economical signals to the distribution utility), not to obtain data 100% error-free, when it is not necessary.

Process Discrepancies

Now, to deal also with a real case, it is analysed a legal charge of “erroneous” processing, in the algorithm to calculate penalties to be refunded to the customers, by the non fulfilment of the quality of supply fixed standards.

A real six months control period is taken as example. The utility refunded 244,632 accounts (10.47% of the total customers), with an average of US$ 2.9 per customer.

In order to audit the results, the Control Authority reprocess again the data from the distribution utility, founding a difference between the calculations. Such difference is of 0.00019% (US$ 1.10 in absolute value), and appears on 361 customers, that is, about US$ 0.003 per customer. It is inconceivably that the utility makes a purposely miscalculation in order to save US$ 1.10 on the US$ 586,341 total, but regardless of this, the legal charge is presented, with the risk of a fine.

These numbers are given in US$ to illustrate the involved amounts (the change rate is nowadays about 3.3 Argentine pesos to a US$), but now it is necessary to swap to Argentine pesos ($), to find an explanation.

When the calculate ENS is negligible, because the customer’s annual energy is very low, or the fined interruption’s duration is in the order of a few minutes, the economical value can be only fractions of a cent.

The utility’s algorithm rounds the economical value to the cent, in each fined interruption (the minimal value that can be refunded), but the Control Authority’s algorithm has been set to a precision of one thousand of an Argentine peso ($).

A customer with two fined interruptions, valued in $ 0.003 and $ 0.002 each one, has no refund by the utility. On the other hand, the Control Authority establishes that the fine is of $ 0.005, and it is rounded to $ 0.01 (US$ 0.003).

Another customer also with two fined interruptions, both valued in $ 0.006 is refunded with $ 0.02 by the utility (each one is rounded to the cent), but the algorithm of the Control Authority ask for only $ 0.01. In this case, there are no legal charges and the calculation is approved, because the Control Authority accepts a discrepancy of $ 0.10 in each individual customer calculation (due to the very complex process).

It is absurd to find that this rule has no application when the discrepancy is between $ 0.00 and $ 0.01 (this fact is not stated in the algorithm official specifications). The exigency is based only on legal grounds and with the argument that the customer cannot loose the right to a refund (even if it is under the practical and attainable level of payment).

The bureaucratic cost associated with this process of legal charges, surpass many times the US$ 1.10 in dispute and it is a waste of time and valuable resources from the utility and the Authority Control. It is an example where not necessary data accuracy is demanded, the clear definition of a process has not been stated, and a theory based legal approach crash with the technical and usual pragmatic realities. The result: a heavy burden on the organisations, and without any profit for anybody. All these problems can be prevented, if before to launch a complex process, all the issues are well defined and a common and clear interpretation as been attained.

IEC TECHNICAL COMMITTEE NO 57 WG 14

The International Electrotechnical Commission standard for a CIM (Common Information Model) is under the responsibility of the Technical Committee No 57. The Work Group 13 deals with EMS (Energy Management System) and the Work Group 14 with DMS [5]. The proposed standard is the IEC 61968, based on the previous EPRI works. Although the standard is not still finished, the present recommendations are a very valuable tool to consider in any DMS project to be developed.

REFERENCES


