RISK ANALYSIS AND AUDIT SOFTWARE TOOLS (DATA MINING)

Alejandro Burr
Director
Internal Revenue Service
(Chile)

Contents: *Summary. I. Risk analysis tools. 1. Introduction. 2. Tax compliance risk. 3. Evolution of risk management in the tax administration. 4. Risk management by segments in a tax administration. 5. Risk area. 6. Analysis tools. 7. The future. II. Software for audits: 1. Background. 2. Definition. 3. General objective. 4. Specific objectives. 5. Product characteristics. 6. SADA audit workflow. 7. Benefits. 8. Results. 9. Conclusion.*

SUMMARY

Within the framework of the present technological development, the role of technology within the tax administrations is fundamental in managing high volumes of information, generated from the taxpayer's behavior.

The great daily volume of data resulting from the taxpayers' obligations has forced tax administrations to modify the strategy and operation for effectively managing their resources.

In this sense, the taxpayers' risk analysis tools play an important role in the detection of unwanted behaviors that otherwise would be very difficult to detect by control officers or could only be discovered once the unwanted behavior has happened.

Also, control support systems - especially those associated to the audit - give many benefits when they are correctly implemented, mainly oriented to standardize procedures, to support resources management, and to leave a registry of officers' actions, that serves as reference for new procedures.

The present document shows how the Internal Tax Service has faced this challenge. On one hand, it describes why it becomes necessary to measure the risk, some examples of the work performed by the Risk Area of the Control Department, and their future challenges.

Finally, part of the experience in the implementation of the Audit Development Support System (SADA), a comprehensive tool for helping the public officer in the control task, will be presented.

I. RISK ANALYSIS TOOLS

1. INTRODUCTION

Evasion and fraud schemes are growing in complexity. The dynamism in their adaptation, modification and transformation, as well as the exponential growth of the information available to tax administrations, added to the fact that taxpayers complete most of their procedures and compliance actions on the web pressure the Tax Administration. Income tax return and VAT from third parties, the taxpayers' own returns for these taxes, the stamp duty and electronic tax documents emission, electronic accounting registries, in the case of micro and small taxpayers (all of the above, in the Chilean case), and many other sources and types of information are now online services. The Tax Administrations have to rely on Tax Intelligence Areas and Risk services to use more effectively and efficiently these new techniques against evasion and fraud, especially when the classic approach requires an intensive previous case-by-case control, through the revision of tax documents, crossing of invoices, expert reports, and many other actions that finally can become unproductive in some cases.

Those new methods and techniques are provided by Data Mining, based on statistics and artificial intelligence, allowing the extraction of useful knowledge, implicit in the available data bases and information sources. Through the models and using data mining techniques, the solution prediction, classification and segmentation problems are approached.

These techniques have been in use for a long time in industries such as Banking and Retail, where they have proved to be effective in risk determination, as for example, the credit risk, and in the application of commercial techniques of profitability and customer loyalty, such as cross-selling (crossed sale).

This document tries to show that these techniques also can be implemented in Tax Administrations and their effects can be measured to analyze their impact in the improvement of compliance levels.
The OECD has defined the risk in Modern Tax Administrations as essential in the definition of innovating strategies to decrease non-compliance levels faced by classic methods.

Chile has been using the experience of foreign tax administrations in risk management, such as the Netherlands, and regional neighbors, Argentina and Ecuador.

Other public institutions, such as the SVS, have been restructured regarding Risk. The private sector has been a pioneer in these matters, especially the Bank and the Retail, that have invested in resources for managing and controlling risks, as essential for creating value for their clients and shareholders.

2. TAX COMPLIANCE RISK

Although the definition of risk according to the Real Spanish Academy is "Contingency or proximity of damage", our Risk Department in the Control Sub-division has adjusted the definition for our institution, as "the probability that a taxpayer does not comply with the tax duties at a given time". This risk definition sets time limits on the noncompliance by the taxpayer with the tax administration, which is directly related to the SII mission.

The risk is one of the most important variables in defining strategies for dealing with taxpayers, in control and audits as well as assistance.

Detecting, modeling, measuring and controlling risks are a task that requires important efforts and resources.

Why do we need to measure and manage risk?

We do it in order to ensure long term compliance and to optimize resource use in higher risk areas.

Risk management

- Risk Management is a formal process where the risk factors, for a particular context, are systematically identified, analyzed, valued, prioritized and attended.
- It is a proactive and systematic analysis of possible events and their answers, more than a mere reaction to these limited detected events. It is the administration of the future.

Strategic risk management process

- To establish the context in which the organization operates, its goals, strategic axis, values and culture and setting a framework for risk decisions making.
- To analyze the organization and environment, identifying the present and potential risks, and:

  – Consider the probability of risk occurrence.
  – Consider the level of reliable probability.
  – Consider the level and reliability of monetary exposure

- Valuing and prioritizing the risks: probability of occurrence, exposure level, cost reduction effectiveness, compared to other resources uses.
- Possible actions on risks, such as: to avoid, to reduce, to transfer, to maintain or to accept them.

Risk valuation in practice

- They are subjective estimates that must be sufficiently reliable, reasonable for the Tax Authority and that could be justified to the taxpayers.
- The risk identification process cannot be very abstract and complex
- Risks may be categorized under a qualitative ranking system, based on historical data, using data mining techniques, or a mixture of both.
- Once risks are prioritized, the risk treatment strategy (avoid, reduce, transfer, maintain) must be defined and resources applied, determining costs in order to select the most appropriate one.
- A residual risk will always exist.

Risk valuation by a tax administration

- Context:
  – To make sure that taxes are paid by those who must do it, for the correct amount and at the correct time.
- Basic Risks:
  – No taxpayer registry; no return; no timely return; no payment.
- Possible causes:
  – Error, ignorance, confusion of requirements, unclear law, economic situation, breaches in the Tax Administration operational processes allowing tax evasion.
  – A high level of tax legal knowledge and assistance are required.
  – The needs of the taxpayers can be detected through a

complex process of assistance and compliance levels can be improved. (Decreasing the noncompliance).

3. EVOLUTION OF RISK MANAGEMENT IN THE TAX ADMINISTRATION

Traditional approach:

- Control each taxpayer.
- Very expensive.
- Same Treatment

Intermediate approach:

- To accept returns.
- To process and review them later.

Risk-centered approach:

- Segmentation of taxpayers into various groups and evaluate those with high default risk and the monetary consequences
- Assistance for voluntary compliance to new taxpayers through education and law simplification.
- Techniques to determine noncompliance, using data crossing:
  – Returns vs. external data
  – Returns vs. industrial data, profession, norms.
  – Returns vs. audits and similar industry results.
  – External data vs. external data.
- These techniques require more documentation and a more formal and standardized risk management process.

4. Risk management by segments in a tax administration

Each segment presents different risk levels => Different approaches are called for:

- Individuals generally are at risk of submitting incorrect deduction of working or living expenses.
- A subset of them, with high income, could perform complex financial and business arrangements to avoid taxation.
- Increased risk in medium and small enterprises segments, due to poorly developed maintenance and records systems.

1. The segmentation must be applied with flexibility and care, especially when business groups have interest in different segments (a specific risk can affect to more than one segment).

2. The tax authority should know the relative compliance ratios, the tax exposure and responsibility of each segment, for locating resources and ensuring compliance.

3. Without a compliance measurement program, the selection of segments to control becomes subjective:

    • Develop sensitivity to the risks and their relative importance according to valuation risk, focused audits, external business information from certain economic sectors and companies.
    • Perform fast verifications of a limited taxpayers group in a specific segment / industry to obtain a risk level.
    • If a mitigated risk is found, a mix of services, such as strengthening the service and administrative changes in the sub-segment can be used.
    • The result of these activities should be evaluated after a period of time to determine at the risk level that still exists and which strategy was the most successful and less expensive.
    • The risk valuation process should be impartial otherwise incorrect decisions can be taken regarding the location of resources and strategies.
    • Separate the risk valuation function from the application of mitigation activities.

5. RISK AREA

The Risk Area has adapted the theoretical methodologies of Data Mining to the institutional reality, in order to generate statistical predictive models (using the term "predictive" for showing that a technique or science has been used, for making conjectures about an unknown fact). This is a process where the requesting unit and / or the one having knowledge of the phenomenon to be studied, beginning with a meeting where the information to use, the potential for solution and their needs are defined. In general, the process includes the following stages (Figure 0 1):

1. Presentation of the modeling proposal.
2. Capturing data to generate modeling variables and target vector.
3. Meeting for Objective Vector agreement.
4. Preprocessing and data transformation.
5. Selection and use of modeling techniques.
6. Determination of patterns.
7.   Interpretation and evaluation of the model.
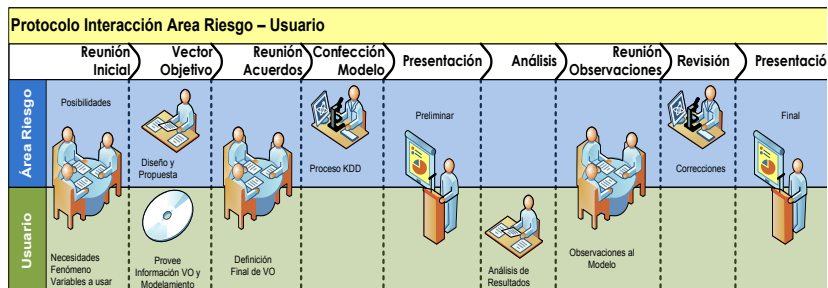8. Transfer of the results to the affected area.

Figure 0-1: review process to establish a risk model.

6. ANALYSIS TOOLS

General models

Estimate models for the general risk, oriented to taxpayer groups segmented by sales size (Large, Medium, Small, Micro Enterprises and Individuals), are called "strategic" because they estimate the probability that a taxpayer does not comply with the correct tax payment. And the use of information extracted from this indicator is not for the selection of taxpayers, but rather to monitor how risk evolves in the taxpayers groups.

To prioritize potential control cases, already selected by a traditional method, overall risk scores, or "strategic" scores are used. For these purposes some algorithms results of "anomalies" are also used.

The risk scores, general or specific, and the "anomalies" can be the income, and another attribute, for building other models that estimate risk.

The risk scores, general or specific, can be used for risks analysis to taxpayer grouped by different criteria; for example the figure shows a matrix of taxpayers by Region (vertical) and Economic Sector (horizontal), and intersections show the average risk score for each group, the most risky ones in red and the least risky ones in green. They can also be used to analyze the risk evolution.

Specific models

Estimate Models for specific risks: For instance "Use of False Invoices". We call them specific since we design to measure the probability of a particular fact. This example estimate the probability that registered taxpayers report their VAT amounts based on false invoices, either from false operations or from not legalized documents.

The risk scores of specific models are used for selecting taxpayers, since they estimate the "control hypothesis".

The risk scores, general or specific, and the "anomalies" can be the income, and another attribute for the building of other models that estimate risk.

The risk scores, general or specific, can be used for risk analysis, for taxpayer groups segmented by different criteria (Figure 0 2). They can also be used to analyze the risk evolution.

Región	Administración Pública	Servicios Financieros	Servicios Personales	Comercio, Restaurantes y Hoteles	Construcción	Transportes	Industria Manufacturera	Agropecuario - Silvícola	Electricidad, Gas y Agua	Minería	Comunicaciones	Pesca	Promedio
Sector Económico													
1	0,00	0,26	0,22	0,06	0,23	0,07	0,31		0,00	0,40	0,18	0,18	0,1
13	0,28	0,10	0,15	0,20	0,25	0,16	0,27	0,37	0,22	0,20	0,28	1,00	0,1
15	0,04	0,15	0,15	0,22	0,21	0,17	0,26	0,32	0,28	0,31	0,32	0,45	0,1
9	0,00	0,19	0,37	0,13	0,21	0,15	0,30	0,30	0,47	0,45	0,69		0,2
2	0,00	0,17	0,13	0,11	0,29	0,43	0,41	0,19	0,47	0,28	0,18	0,20	0,2
4		0,25	0,52	0,11	0,08	0,19	0,20	0,49	0,46	0,45		0,67	0,2
8	0,00	0,19	0,28	0,19	0,25	0,23	0,35	0,37	0,49	0,53	0,19	0,45	0,2
14	0,22	0,22	0,15	0,24	0,28	0,30	0,28	0,35	0,29	0,43	0,61	0,70	0,2
5	0,00	0,26	0,31	0,19	0,24	0,30	0,30	0,40	0,38	0,22			0,2
19		0,29	0,23	0,09	0,23	0,20	0,30	0,45	0,33		0,50	0,50	0,2
7		0,18	0,35	0,22	0,33	0,26	0,31	0,40	0,40	0,47		1,00	0,2
3		0,21	0,24	0,19	0,36	0,22	0,13	0,39	0,21	0,53	0,44	0,40	0,3
12	0,00	0,19	0,17	0,20	0,52	0,27	0,52	0,44	0,21	0,51	0,75	0,65	0,3
17	0,53	0,57	0,53	0,50	0,58	0,46	0,64	0,63	0,60	0,72	0,69	0,78	0,5
Promedio	0,16	0,19	0,22	0,23	0,26	0,27	0,37	0,40	0,43	0,47	0,48	0,60	

Figure 0 2: taxpayers matrix by region (vertical axis) and Economic Sector (horizontal), at intersections is the average risk score for each group, the most risky in red and the less risky in in green.

Anomalies

The "Anomalies" are algorithms used to determine odd behavior from of a business or tax point of view, which not necessarily means noncompliance. For example, taxpayers who do not submit their monthly returns but keep requesting the legalization of tax documents.

To prioritize potential audit already selected by traditional methods, overall risk scores or "strategic" risk scores are used, as well as some results from of "anomalies" algorithms.

The risk scores, general or specific, and the "anomalies" can be the income, and another attribute for building other methods that estimate risks.

Measurement methodologies

Measurement methodology is associated with the development of an observation method for auditing the taxpayers' future behaviors. This method uses as a comparison basis the past behavior of the taxpayer or with his segment or sector, among other analysis relevant for control purposes.

7. THE FUTURE

Quality of information evaluation

The tax administrations work is based primarily on managing accounting, economic and financial information of taxpayers, so the quality of the information that enters the various analytical processes is essential; and additional measure in this sense is the design of a methodology, using statistical criteria, that will estimate the quality of the information entered in our databases, and to define strategies and actions that improve the data quality.

Text mining

The Risk Area team is preparing to enter a new dimension of analysis information, in this case unstructured data, acquiring a text mining tool, and training in techniques to usefully exploit this type of analysis. The text mining methodology allows performing analysis of open texts containing accounting, economic and financial information from taxpayers, for example the product description on invoices and financial reports, information from web pages and other publications, etc. In short, there are many possibilities.

Web Mining

In view of the current challenges and due to changes in the way we do business, it is essential to strengthen our business intelligence service with information collection actions and monitoring on the Web, in regard to activities associated with taxpayers, industrial sectors, products or services, in which potential commercial transactions are detected related to the marketing of products or services, that given the current forms of control, need specific tools for this purpose

II. SOFTWARE FOR AUDITS

1. BACKGROUND

When officials perform a tax audit process, documents verification, and VAT refunds, among others, in addition to performing the control, they must spend time for the administration and registration of the case development.

In this area, a few years ago the Internal Revenue Service did not have a tool that would allow tracking the development of a tax audit, and also solve control process management, as well as having an automatic registry of events associated with tax audits made to taxpayers.

The purpose of a tax audit is to verify the taxpayers' compliance. Therefore, tax returns correspond to the accounted operations according to their commercial value, to the tax and accounting support documentation, and reflect all transactions or operations performed.

First, audits are assigned and monthly distributed to officials of each office in the country. This procedure used to be performed through email or paper. Thus, during the control process which took place in offices, generated the required documents to be delivered to the taxpayer through notification; which the taxpayer had to review in order to clarify certain differences on taxes. The documentation submitted included the payments for the settlement. The TA requested a payment of the tax differences detected. Among the actions from an audit, the possibility that the taxpayer acknowledges the tax differences and correct his statements was also considered, it was also feasible that the taxpayer fully clarify the information provided, thus ending the control. In any case, during the tax audit, the documentation provided or removed by the taxpayer was registered, using documental receipts or returns.

To complete each of the above mentioned forms, the auditor had to enter in the applications the information required to perform the actions (name / business name, address, legal representatives, and others specific taxpayer data). In each of these actions, a number of subtasks were identified; recording every event in the audit, but these tasks increased the workload in each of the cases under examination.

Certainly, one of the complex instances within the audit consists in establishing the legal background that underlies the tax differences determined in the audit. Such legal grounds play a key role in those cases in which the taxpayer makes use of his rights under the tax law,

particularly when he appeals against actions taken during the audit, to the tax Courts, which currently operate autonomously.

As a result, the control process resulted in a set of worksheets available to auditors, both for the internal audits and for the taxpayer registries, which eventually turned into case folders. In addition, during audits, manual notes were taken for the Tax Administration, reporting the taxpayer's situation, which in some cases allowed controlling the business operations, through denying the legalization of tax documents.

It should be mentioned that when taking these notes it was always possible to make an error in the procedure error, and the appropriate entries regarding the taxpayer noncompliance were not always entered in the systems.

The method used to solve this problem was the implementation of a tool based on information technology that would monitor the workflow related to the audit performed to taxpayers, in this sense, building such tool provides the necessary support to officials for their specific audit tasks.

Indeed, the completion of this project allowed facilitating and supporting the audit by the tax officials, for standardizing these actions, and having permanent, accurate and timely information related to the management of national and regional tax audit.

2. DEFINITION

The Support Audits System (SADA) is a technology platform that manages the entire business flow of tax audit.

This platform includes tools that enable the Business Process Management (BPM), in which all workflows associated with an audit are managed.

Therefore, the platform integrates processes modeled, with web services that provide the internal information to the institution regarding the taxpayer, thereby increasing the information available to officials, the value of the information technology infrastructure and optimizing the use of the workforce.

This working environment also facilitates and supports the documentation of the activities developed in the control process.

3. GENERAL OBJECTIVE

Implementing SADA nationwide intends to standardize the procedure to carry out a tax audit, providing the officer with a software tool and a work environment in order to facilitate the monitoring and the generation of audit tasks.

4. SPECIFIC OBJECTIVES

Implementing the system:
- Facilitate monitoring and management control processes from the Program Registration until the financial transfer, allowing the generation of information for the workload management and administration.
- Facilitate standardization in execution and registration of actions performed by the Service through management and generation of various tax audit processes
- Keep an automatic record of cases history (as the officer develops the case using the tool), which will allow to have accurate and timely case management information.

5. PRODUCT CHARACTERISTICS

The project included the development of the following features:
- Registration of control programs: Allow registering programs designed for control areas, including their objectives, taxes, taxpayers, periods and documents under review, and any relevant information within the programs.
- Case entry, selective and emerging case assignment control: It is the control and monitoring of the cases that are included in centralized audit plans.
- Entry of emergent cases: It is the control and monitoring of cases arising out of the plans, which are generated by a same unit. They allow making a special plan for the case to be analyzed as well as the assignment of an inspector responsible for the audit, and follow the other audits procedures.
- Audits Follow-up: Include the whole process of taxpayers audit (Notifications, Citations, Settlements, etc.), according to the plans provisions, allowing to keep track of all the activities and tasks involved. This module incorporates the history of taxpayers not required and requested, open actions such as notification, settlement, provisional settlement, reassessment, return of documents and registry of the case conclusion.
- Entry and follow up of refunds: Allow registering the taxpayer's request for refund and the result of this request, generating legal documents, automatic or manual resolution depending on the type of tax. Here the VAT refunds for purchases of fixed assets (27 bis) are registered and VAT for subject change.

- Monitoring of document verification: allows recording the documents to review, entering the verifications for each case and the results of this process, allowing repository of verified documents. It also provides information on the revised documents.
- Reports: allows an overview of the tax audits assigned to an office, so the management can take decisions regarding their development.

6. SADA AUDIT WORKFLOW

The process begins with the entry of the centralized cases load, and their automatic distribution to each of the country units (Plan); the SADA automatically captures information from other Tax Administration applications. The units can autonomously enter cases (emerging) to monitor, if they could represent tax evasion schemes.

Next, the workflow starts generating the corresponding tasks, so that officials can proceed to the specific actions (Case Notification, Rectification, Citation, Settlement, and Case Termination). Thus, with the information entered into each of the screens that compose each action, the system will automatically generate the necessary documents to communicate formally with the taxpayers; it should be noted that these documents have replaced the paper forms previously associated with each stage, which were issued independently by each service office; this has facilitated the standardization of procedures at country level (Figure 0 3). It should be noted that the process of verification of documents and request for refunds are actions controlled by the tax authorities on certain operations of the taxpayer, and could constitute an audit.
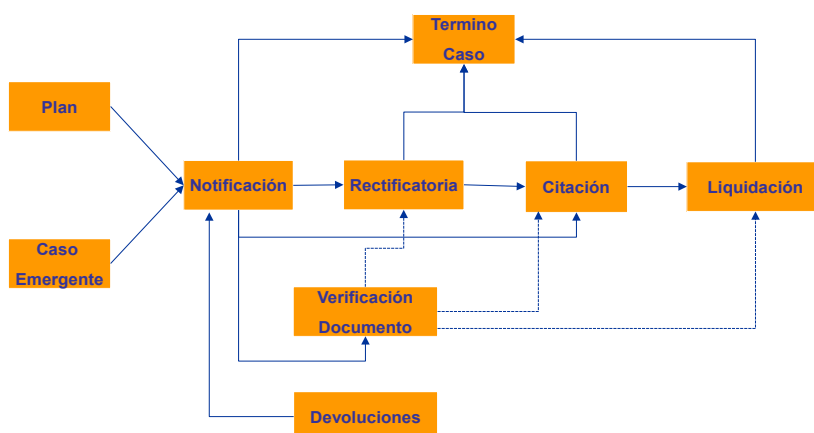

Figure 0-2: SADA Audit flow chart.

The workflow allows, for each stage, to generate and automatically upload information through Web services for the tax administration database. This keeps updated the information related to each taxpayer control, and the notes that previously were done manually are now automatically loaded in the corresponding internal systems. Similarly, the system collects the taxpayer information from other systems and databases to generate actions, related primarily to the taxpayer's own information, tax behavior, and forms submission, among others. Officials do not have to access the systems to update this information; this reduces the time spent in auditing for completing repetitive tasks, minimizing control process time.

In order to facilitate the officers work in the control process and to ensure a better legal basis for each of the tax differences detected from the tax audit, it is to note that instructions associated with the use of the system highlight the importance of standardizing the various figures of tax evasion; in this sense, the system provides the officer with a basic structure for each one of the detected differences, through the selection of the legal framework needed to explain and support these differences.

7. BENEFITS

For auditors

- **Simplification of the daily control work, by accessing automatic information from other applications (taxpayers' registry, monthly and annual tax forms):** Via web services, the application communicates with other internal systems which have the necessary information to create the control documentation; this allow determining the amounts the taxpayer has to pay.

- **To report pending cases and their history***:* The program includes an user menu which allows clarifying the pending tasks for audit, and store the information, which can be consulted at any time.

- **To allow the creation and registry of the performed actions:** The system allow saving the information entered for any action, and also generating the document which represent the action requested by the controller (certificates of settlement, notices, summons, assessments, reassessments and provisional assessments).

For authorities (Group, Department and regional Director)

- **To have a system that supports the administration of cases and tasks**: The system enters the cases that are distributed to the units, and through the workflow, assign them to the auditors in charge, which allows for clarity of the workload in each unit.

- **To be informed on the development and outcome of the cases**: The system has modules for consultations according to the profiles. This allows access to each case parameters; for the cases that are completed, it is possible to know the associated yields, and for pending cases to view the tasks already carried out.

- **To standardize performance reports:** The workflow creates tasks within a performance (notices, summons, settlements, etc.) Through a logical order, structuring the way the review process is done, ensuring the quality of the action.

- **Reporting:** The system allows exporting data to other formats and generating reports with the information that the user requires.

For the national audit management

- **Time optimization in the control process:** The automatized system allows a better efficiency and better use of the controllers' time for control activities, most of all because the review deadlines are regulated.

- **Standardization of the quality of the settlement:**   The legal support provided by the application, through a justification of the detected differences allows backing the observations transmitted to the taxpayer. In case of appeal to administrative courts, this allows a better follow-up and control.

- **Reliable information is available online:** Allow timely information about the tax control actions registered in the application:

  – Workload information.
  – Cases yields.
  – Cases history registry.

- **Feedback in control processes:** The system allows determining the controls yields, and associated concepts, redefining the control programs or creating new ones in cases where it is necessary if important elements of the case were not mentioned

- **Standardization:** This allows that the processes involved in the audit are carried out homogeneously throughout the tax administration, independently of the particularities that may exist in each office.

- **Integration with other applications:** The option to integrate SADA with other internal applications not only allow transferring information from one system to another, but also allow the complete result of the audit, including what the taxpayer has transferred, and what was entered in the treasury.

8. RESULTS

The use of SADA began nationally in 2012. For this period, the amount collected from audits by the Internal Revenue Service was $ 1.356 billion, which is obtained from Selective, Emerging and Massive Examination actions. By breaking the previous figure into selective and emerging control concept, the total registered cases in the system is approx. 5,800, with a yield of $ 735 thousand million (Table 0 1).

	Amount of cases audited	Revenue collected (MM$)
Selective audits	3.664	634.421
Emerging audits	2.188	101.310
TOTAL	**5.852**	735.731

Table 0-1: cases audited and revenue obtained via SADA application in 2012; the performance of these cases represents 54% of total revenue from control.

9. CONCLUSION

The Audit Support System (SADA) is a management tool that has allowed enhancing the control process. It is the application where all audits independently from their origin will be developed, since its main advantage is to standardize the process, along with retaining all information related to the cases, which is useful when generating reports to show the case status, performance, development time, among others; it also contributes to resource management, as well as to measure the efficiency and effectiveness of the audit work. The SADA is a flexible application, making possible to implement new tax arrangements applying a continuous improvement in the audit process.

Through the use of this tool, the audit manual process is replaced by an automatic one, which eliminates errors inherent to manual repetitive tasks. It should be noted that the analysis of each case by an officer during audits remains the most important raw material in the process, so the tool only intends to help carrying out all the associated tasks.