

Audit Strategies in a Great Data Base Context: the SecexPrevidência Experience



Fábio Henrique Granja e Barros

is a civil servant of the Federal Court of Accounts - Brazil. He is the Head of the Department of Social Security, Labor and Social Assistance Audit (SecexPrevidência).



Melchior Sawaya Neto

is a civil servant of the Federal Court of Accounts - Brazil. He is the Director of Social Assistance of the Department of Social Security, Labor and Social Assistance Audit.



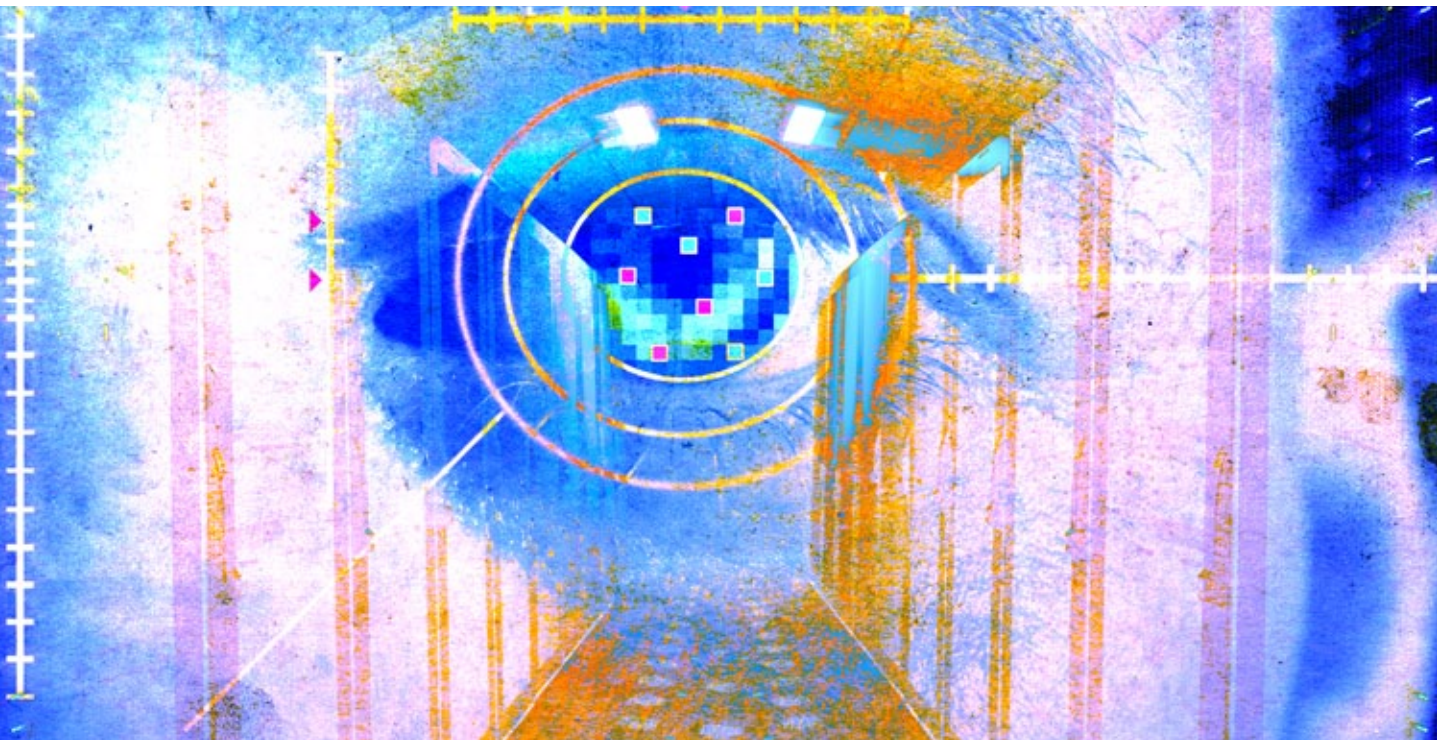
Jorge Mendes de Oliveira Castro Neto

is a civil servant of the Federal Court of Accounts - Brazil. He is the Director of Social Security of the Department of Social Security, Labor and Social Assistance Audit.



Teonio Wellington Martins

is a civil servant of the Federal Court of Accounts - Brazil. He is a Director at the Department of Social Security, Labor and Social Assistance Audit, performing at the department responsible for the oversight of the Ministry of Labor and Employment, of the "S" System institutions and of some Professionals Oversight Boards.



ABSTRACT

The nature and volume of the expenses, as well as the amount of available data facing the government functions related to work, social assistance and social security, imposed upon Secex Previdência (Department of Social Security, Labor and Social Assistance Audits) the need to employ a data analysis in order to improve the effectiveness of their audits and evaluations. For that purpose, public and private basis crossing and analysis of quantitative efficiency, governance and of the impact of social expenses in inclusive growth were performed. The Secex Previdência implemented, with a software enterprise, an analytical board to test a methodology that could integrate several quantitative analyses for the generation of better diagnosis, prescriptions and predictions for the identified problems in order to improve the currently used analyses. In this sense, it was possible to test the enterprise's IT solutions in the integration of the results of several techniques of data analysis for two benefits operated by the National Institute of Social Security, age benefits and for people with disabilities.

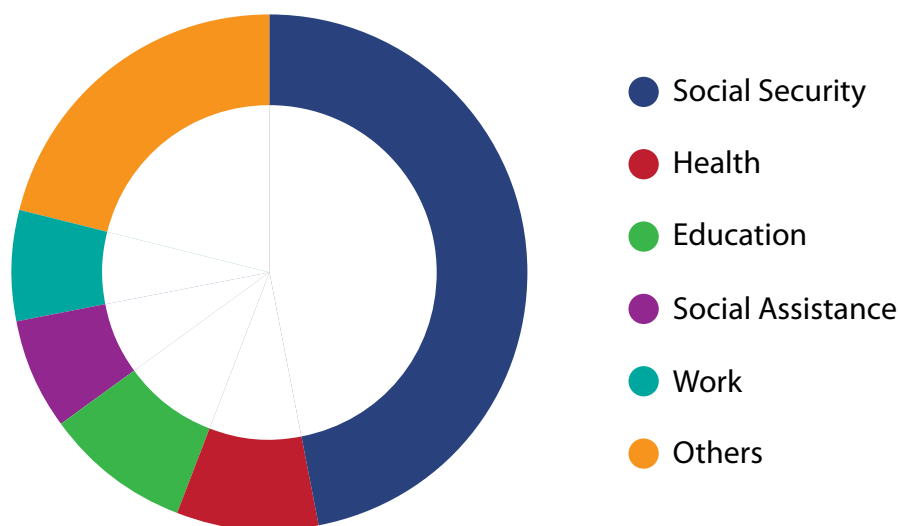
Key words: Data analysis, quantitative methods, IT, Social Security, Assistance, Work

1. INTRODUCTION

SecexPrevidência performs in Social Security, Social Assistance and Labor areas. These areas represent 61% of the Federal Government Budget (graph below), not including in this figure the expenses with the public debt duty. Retirement, pensions, Unemployment Insurance, Special Salary Raise, the Bolsa Família Programme, Continuous Benefit Conveyance, agreements and transferences from one fund to another are among the main expenses monitored by the Audit Department.

Due to the relevant amount of information that Secex Previdência monitors (over 50 million of beneficiaries, 900 Social Assistance Reference Centers, 1500 Social Security Agencies, 1200 Sine Agencies, database with a great amount of records (MACIÇA, CNIS, RAIS, SISOB, Folha PBE, CadÚnico, CAGED, SIAPE, CensoSUAS))¹, the development of several strategies using quantitative methods and data analyses were necessary.

Among the Audit Department's experiences, the Audit and Oversight Strategy of the Social Security Benefits that comprised the analysis of internal controls and the implementation of three audits in the databases on the benefits of greater materiality (rural retirement, death pension, retirement due to age or contribution times) was mentioned. In addition to the review of possible irregular benefits, these audits pro-

Graphic 1:Expenses engaged -
by function (2013)

Source: Siga Brasil

vided the strengthening of concession systems, once the scripts from the database crossing were transferred to the NISS, enabling the body to apply more efficient controls in over 30 million active benefits, reducing the probability of errors and frauds. The cessation of the benefits granted is equivalent to a potential economy estimated at approximately R\$140 billion per year (Ruling 456/2010-P, Ruling 715/2012-P, Ruling 666/2013-P).

Similar to this strategy, in the Board of Works, data audits have been performed since 2012, in the Unemployment Insurance on the formal worker and artisan fishermen modalities. The Unemployment Insurance represented an approximate expense of R\$32 billion in 2013 and it benefited 5 million people. The information of the administrative record of the Unemployment Insurance were crossed with the following public database in those audits: CPF, RAIS, CAGED, CNIS, SIAPE, MACIÇA, SISOB, TSE (Supreme Electoral Court) candidates and Bolsa Família). Millions of instalments of the Unemployment Insurance identified were irregularly paid, for that reason the recovery of the resources and the improvement of control systems were determined, estimating a potential control benefit of approximately R\$150 million (Ruling 2.089/2013),

Efficiency analyses were also performed in the agencies of Sine, Social Assistance and Social Security. For that purpose, a quantitative approach was sought for the efficiency measurement, the Data Envelopment Analysis – DEA. The technique enables the quantification of the relative efficiency level between opposites, showing the lines of productivity improvement and benchmarking units.

The DEA has its use disseminated in several economic sectors, such as: electric power, telephony, ports, highways, schools, hospitals, among others. The technique enables the efficiency forecast in administrative units, once it builds a production frontier comprising efficient units, which are compared to inefficient units. The efficiency comparison, performed with the technique, is always relative. In other words, some units are considered inefficient because there is a more efficient unit with a similar production profile. In addition to demonstrating the feasibility of improving the efficiency, the relativity presented by the technique also shows the possibility of improving efficiency for more effective units that compose the production frontier.

In relation to the results, from the 1,207 agencies of Sine, 1,500 Social Security agencies and 6,893 Social Assistance Reference Centers, only 19, 23 and 406 are on the efficiency frontier, respectively. The results of these efficiency analyses were very promising, for the possibility of enhancing the amount of attendance was noticed in all three functions of the government (Social Security, Social Assistance and Work), which means services to citizens with no need of raising costs.

Another quantitative analysis performed using the National Household Sample Survey (PNAD) of the Brazilian Institute of Geography and Statistics (IBGE) was the evaluation of the impact of the social security on the inclusive growth (Ruling 1274/2013 – Plenary). The analyses performed indicated that the rural subsystem of the Social Security General Regime (RGPS) performs a relevant social inclusion function, as: 1) it

generates an income increase proportionally larger in household with lower per capita income; and 2) proportionally forwards more resources in relation to the GDP to municipalities that present the lowest GDP per capita. This report was presented in the 2012 Government Accountability.

A quantitative technique named “main components analysis” was also performed using the CEN-SOSuas to define a governance rate for the Municipal Boards of Social Assistance. The action had the participation of a sociometric expert from the Federal University of Brasília (UnB) and it enabled an indication of the capacity of those boards to operate and audit the resources transferred from one fund to another. This result is expected to be used as a definition for future audits and to evaluate the social control that has been performed by those institutions in a more accurate way.

Although the data alignment methodologies and the quantitative methods currently applied by SecexPrevidência involve important results, it was understood that it would be necessary to develop a methodology able to integrate several analyses in order to generate better diagnosis, prescriptions and predictions. For that purpose, an analytical board was performed with a software company, from February to June 2014, aiming at testing the enterprise’s IT solutions on the integration and automation of the results of several data analysis techniques for two benefits operated by the NISS, by age and for people with disabilities.

Therefore, the following benefit types were selected: retirement due to age and Continuous Benefit Conveyance for people with disabilities. In a general way, the joint collaboration with the software enterprise consisted on the cleaning and integration of used databases, on the identification of typologies that characterize irregularities, on the use of several data mining techniques that highlight fraud patterns and abnormality patterns and, finally, on the integration of these results in the elaboration of a metric indicating the probability of irregular benefit.

In relation to the cleaning and basis integration, a Data Quality was performed, which enabled the identification of the Issuance of Individual Taxpayer Registry of 99.2% of all beneficiaries identified on the MACIÇA of February 2014 with the probability of over 95%. Phonetic identification of match-code generation and record duplication techniques were used, resulting in the identification of the beneficiary’s CPF through the analysis of the data contained in the

MACIÇA and in the CPF of the Brazilian Federal Revenue (RFB), considering the content of the following fields: CPF number, name, birth date, mother’s name, voter ID card and home address.

Several data analysis techniques were used, from which the following must be highlighted:

1. **BublePlot:** the technique was used to analyze the amount of Unemployment Guarantee Fund (FGTS) payment forms and Social Security Information (GFIPs) presented by and employer in each month. The standard is that an employer presents only one statement of employment relationship through the GFIP system each month. Occasionally, for the amendment of any error or omission, the employer may present a second statement. The objective was to highlight the cases considered abnormal, different from the standard and with a great number of amended statement.
2. **Logistic Regression:** the technique was used to detect the cases considered abnormal, different from the standards of concession of benefits (in this case, only two types of benefits were analyzed, which were the scope of the analytical board).
3. **Relationship networks:** the technique was used to analyze the existing relationships between the data of benefits that were proved irregular and the data of benefits paid in February 2014. The analyzed relationships refer to benefits of a same holder, settlor, recipient or prosecutor. In other words, all benefits that are being paid and present relationships of that nature with benefits that are proved irregular or that were selected as suspects or demanding a more detailed analysis.

Finally, the consolidation of the results from all analyses was performed using IT solutions from a specific enterprise and it enabled the score definition for each benefit. In other words, some sort of a score-card matrix was used. Numbers and colors according to its value represented the score. The highest values (a maximum of 1000) indicate a high probability of an irregular benefit.

It is understood that this methodology will enable SecexPrevidência to perform in a selective way

on the benefits with greater possibility of error/fraud, on the vulnerability of transactional systems, on the decrease of false-positives, on the monitoring of deliberations, on the performance of more accurate diagnosis, on the identification of opportunities that may contribute to the delivery of best services to citizens. The methodology will also enable the automation of some data-crossing procedures, which may be periodically performed at a low operational cost. In short, the implementation of this type of methodology will contribute to the improvement of techniques of fraud detection, risks evaluation and transparency dissemination currently used in the Audit Department.

Given the materiality and complex of social security systems regimes, benefits of Unemployment Insurance and Special Salary Raise of the Ministry of Labor, and of the actions of income transference of the Ministry of Social Development and Fight against Hunger, the need of implementing actions involving the use and analysis of a great database is reinforced.

Therefore, the challenges faced by SecexPrevidência in this action strategy involving data analysis in an integrated way require changes in the work processes of the unit as well as in the articulation with other areas of the Court, especially in the area responsible for the training and information technology.

In relation to the support to the IT area, a new way of integration among and within teams of the Court is necessary, evolving to the mass use of IT in the integration of bases, quality increase and development of solid and useful analyses for the responsible areas of the Court. That requires training, integration among teams of different areas and an adequate software for data mining for information crossing.

The conjunction of several actions is necessary to achieve the desired performance scenario. The first and most important is training auditors that already perform in the SecexPrevidência in areas of new challenges knowledge. In addition, it is proposed that a group of auditors from the Audit Department become responsible for the interaction with other areas, especially from the IT, and for the information support that may become necessary for control actions of the unit. This group would aim at contributing, through analyses and database alignment, to a greater effectiveness and efficiency of the resources allocated in the Audit Department while guiding the actions to benefits and transactional systems that present a greater risk. The action would improve the quality and the results of the works, for a diagnosis performed in 2014, supported

by Seplan, identified a significant and repetitive investment of time on data analyses that could be part of a previously structure routine, providing the auditors with more accurate and continuous information bases.

However, based on a diagnosis previously performed, it was identified that the implementation of a strategy in SecexPrevidência that enables the use of data in a broad manner may increase efficiency in the performance. For that purpose, a modification in the processes of internal works, of personnel allocation, IT support and technicians training will be necessary.

In addition to present the systematic of integrated analysis in a context of great databases being developed in SecexPrevidência, the present article aims at drawing attention to the opportunity of developing similar strategies in other areas of the Court, as in Health, Education, Personnel, Works and Bids, once they all work with broad databases.

NOTE

- 1 MACIÇA – payroll from the benefits operated by the NISS; CNIS – National Registry of Social Information– registration of those covered by the NISS; CadÚnico – Social Information Registration – registration of social assistance beneficiaries of the Ministry of Social Development (MDS); CAGED – General Employed and Unemployed Register; Folha PBF – Bolsa Família Programme Payroll – under the responsibility of the MDS; RAIS – Annual Listing of Employees and Salaries from 2009 to 2012; SIAPE – Human Resources Management Integrated System; CPF – Issuance of Individual Taxpayer Registry; SISOBI – Obituary Control System of the Ministry of Social Security, CENSOSuas – Single System of Social Assistance census.