

Preventing and Detecting Fraud through Data Analytics in auto insurance field

Authors:

Adrian BĂNĂRESCU*

Aurel-Mihail BĂLOI**

***Abstract:** Currently, in the autoinsurance field, fraud has become a real phenomenon, and insurance companies are making efforts to identify and implement effective means of fraud prevention and detection. This article aims to present the data analysis and related tools, as a possible alternative to fraud, in the context of the current information society (I-society). In the literature, data analysis processes are presented as a highly effective means for the prevention and detection of fraud, but their implementation at both inter and intra-company levels is a difficult process, because it involves organizational transformations in which the decision should be based on knowledge. Although in Romania the institutional framework for fraud detection in the auto insurance field is created, the information processes that take advantage of the potential of intangible assets (data, information, knowledge, relationships, etc.) are not systematically exploited and are not integrated into decision-making (whether strategic or operational).*

***Keywords:** fraud; IT Systems; data analytics; network analysis; geospatial analysis*

***JEL Classification:** G32, K42, O39*

* Adrian BĂNĂRESCU, Postdoctoral researcher at the Institute of National Economy, Romanian Academy, e-mail: rescuadrian@yahoo.com

** Aurel-Mihail BĂLOI, Intergraph Computer Services, Romania, e-mail: aurel.baloi@ingr.ro

Introduction

Human society is in a process of continuous transformation, being marked, at least in the last 20 years, by profound mutations produced by the content change and knowledge development. Technological revolution provides opportunities for development in all fields of economic and social life, being, in the end, an intrinsic component of our social life. There is a behavioral discrepancy in assimilation of technology by organizations and individuals, sometimes the latter adopting technologies much faster than organizations. In Romania, this phenomenon is on the increase, gaining items of specificity, due to financial constraints occurring within the current economic environment, dominated by companies that have a turnover below 1 million euros.

It is a well-known fact that currently nothing remains constant so, profound changes, radical innovations are made and are implemented in a very short period of time. What is even more impressive is the frequency of these changes and the consequences they produce at economic and social level (*large companies on the telecommunications market promote a new phone model at every 6 months and the life cycle of a car, decreased in the last decade to maximum 5-6 years*). All this technology provides significant amounts of data on an ongoing basis, and their analysis by means of advanced IT systems allows for objective decisions in real time. Among individuals, those who take advantage of technological opportunities on very short notice are the criminals, and auto insurance is not deprived of such behavior.

Partial completion of literature shows an increasing trend regarding the tendency to use modern analytical methods for detecting fraud, mainly in areas such as finance, banking, insurance or public order and national security.

In a broad sense, we can say that the data analytics help companies to make more informed antifraud decisions, quantify the degree of compliance with regulations and procedures, creating the possibility of identifying irregularities, inconsistencies, determine operating modes, and predict certain activities. Analytical products drawn from data analysis can provide premises for ensuring conditions for fraud detection and quantification at the level of an economic entity.

On the grounds that insurance companies generally operate with large amounts of data, we propose concrete solutions to implement recovery processes, in order to identify anomalies in the data stream or potentially fraudulent behavior patterns. The information thus obtained will be used later in directing investigations, and to make recommendations aimed at improving control activities.

We intend to highlight the importance of the implementation, development and operationalization of an *early warning system* for prevention and detection of fraud. In terms of cost, it is recommended to improve fraud prevention in the context of a low degree of recovery of damages caused by illegal activities (*the causes are multiple, including the possibility of circumventing the seizure of assets*). Fraud prevention can be achieved by deploying a system of risk management and a warning process, "red flag"/"hot spot", built on a set of indicators which, being unusual, trigger the prerequisites checks.

In order to counterbalance the slow pace of economic recovery in the context of the permanent changing threats, managers should set up a series of anti-fraud measures, as a lever to control the costs, in terms of reducing available resources, which implies an active and comprehensive approach.

1. Insurance fraud issues

The fact that we are witnessing an increase of the insurance fraud phenomenon is also confirmed by the recent amendments to the criminal code. In support of the work of counteracting illicit behaviors in the insurance field, the legislator has incriminated this type of fraud, as a distinct offence.

Assuming that the rules of incrimination should relate to the needs of society at a certain moment, the new criminalization of insurance fraud was based on a social reality, being known that, especially in times of crisis, such facts have a higher incidence.

From another perspective, the choice of the Parliament to create a separate indictment that relate to the deception of insurance is justified by the fact that the efficiency, in terms of general prevention, is higher to the extent that there is a special dedicated regulatory (Miron, C., 2012).

Auto insurance fraud is particularly favored by the ease with which it is committed, as well as the huge amounts that may result from settlement of a case on damages. The simulation of accidents is among the most common *modus operandi* along with overvalued invoices that are paid off.

Measuring the level of auto insurance fraud may be considered a challenge today, as there is a single authority with specific powers in this respect. Although insurance companies carry out their own analyses, they are reluctant to formalize the statistical data relating to fraud, for reasons of domestic policy and marketing image.

Frequently, the internal frauds are not reported by companies, which can lead to loss in credibility on the part of their clients. At the same time, the fraudulent internal operations damage the reputation of the entity affected by this scourge. On the other hand, cases are not investigated systematically, so that it is difficult to count the frequency of fraud crimes or the amount of damages caused.

Although there are no statistics on the extent of the fraud insurance in Romania, a slight increase in attempted fraud as a result of the economic crisis may be noticed. Moreover, specialists predicted that in 2012 insurers would have registered a loss of about 100 million EUR as a result of fraudulent claims for damages. According to a survey conducted by 1asig.ro, more than half of respondents believe that fraud accounts for over 20% of total compensation, other 33.3% estimating fraud between 10% and 20% of the value of claims paid by insurers.

In the presented context, the differentiation between competitors operating on the market in Romania, will be made by their ability to analyze data and make decisions fully aware and in real-time. Processing and storage resources always endorse data analysis, thus making it possible to transform the relatively uncorrelated data in important data that may determine the future of a company.

Public data shows that the main institutional actors in the insurance field take steps for setting up and developing a system of prevention and detection in this field, the main initiatives being the following: the creation of a common database to verify the vehicle damage history ¹; use and development of electronic filing applications of auto-events registered by vehicles; improve internal control procedures, antifraud internal procedures and prevent and combat money laundering and terrorist financing.

Regardless of size, market share and company reputation among insurers, the main challenges are: the specificity of the current financial crisis that has brought to the fore that fraud occurs mainly during a recession compared to normal periods of economic growth; precautions in the choice of lowering customers' fierce competition in the market and the desire to achieve the targets set by investors; permissive or ambiguous legislation; lack of staff and investigative tools; increasing periods of investigation and hence the identification of fraud, because of specialization of fraudulent subjects; traditional methods of fraud identification have certain limitations.

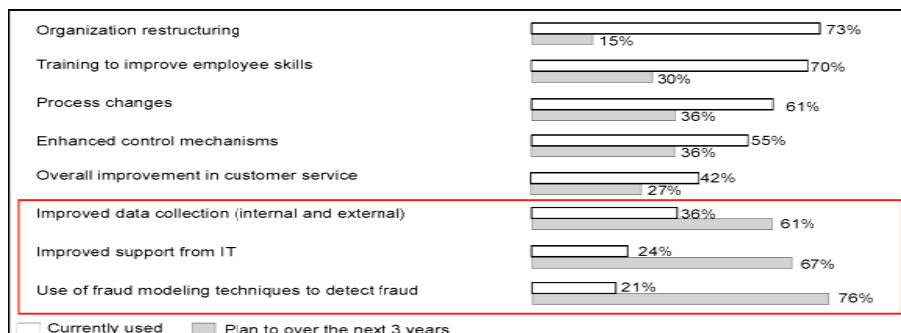
¹ <http://www.dailybusiness.ro/stiri-finante-banci/ce-vor-face-asiguratorii-pentru-a-nu-mai-fi-fraudati-pe-segmentul-casco-100120/>

On the other hand, the need to protect insurance companies encounters the need to protect personal data of the insured person. Some even propose legislative initiatives to rule their insured right "to be forgotten" and to withdraw consent to the processing of personal data by the company. From such a perspective, the fraud detection system would become obsolete by emptying the content databases and thus obliterate indications of fraud.

From the perspective of data analysis, we consider that the claimant for damages speculates that insurers do not benefit from the expanded capabilities of predictive or geospatial data analysis, data mining or social network analysis.

As regards these problems, we believe that the issue of data analytics is a key factor for Romanian insurance field. According to Accenture 2013 Europe and Latin America Claims Survey, the respondents see an attractive opportunity for investing in claims fraud analytics and other anti-fraud solutions. Respondents also believe the most important initiatives for strengthening their anti-fraud capabilities would be to use fraud modeling techniques (52 percent claim this is critical), improve the collection of both internal and external data (37 percent), and obtain greater IT support (31 percent). Looking ahead, the major focus is on organizational restructuring, training and process improvement to help strengthen fraud detection. However, the priority on short to medium term is to use fraud modeling techniques (76 percent), improve IT support (67 percent) and improve data collection (61 percent).

Figure 1 – Key Initiatives Implemented to Improve Fraud Detection



Source: <http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture-How-Effectively-Fight-Insurance-Fraud.pdf>

2. Reasons for using data analytics and information processes

Annually, global organizations declared losses of billions of dollars because of many cases of fraud. This phenomenon has multiple implications, manifest at any level, and the effects are the most diverse, starting for the decrease of the profit, affecting profitability, diminishing consumer confidence by damage brand image. The stated aim of any fraudulent agent is to avoid fraud detection means, meaning the operating mode will play a key role in achieving this goal, the methods of fraud are becoming more sophisticated.

The economic impact of fraud results in increased costs generated by fictitious or inappropriate processes and services and the amounts diverted, which will no longer be part of the mass of income or profit.

During fraud investigations, the data should be evaluated objectively, on the basis of various analytical methods and techniques, in order to formulate hypotheses and recommendations for making a decision. In the presented context, we consider that operationalization of a data analysis system could lead to the reform of the current control structures of the companies, which will benefit both the modern means of warning as well as of support during the investigation. The purpose of applying of methods and techniques for detection of fraud consists in creating an optimal environment of work, at any level, designed to enable the organization to gain control over expenditure, to minimize the risk, to limit fraud and thus to increase effectiveness.

There is a spectrum of analysis that can be used to detect fraud. It ranges from point-in-time analysis conducted in an *ad hoc* context for one-off fraud investigation or exploration, through to repetitive analysis of business processes where fraudulent activity is likely to more likely to occur. Ultimately, where the risk of fraud is high and the likelihood is as well, organizations can employ an “always on” or continuous approach to fraud detection – especially in those areas where preventative controls are not possible or effective¹.

In The Global Economic Crime 2014 Survey, Pricewaterhouse Cooper (PwC) specialists decide that data analysis is a modern method of identification of critical fraud. However, this method has proven effective in early detection of fraud, accounting for 9% globally, 4% in USA and 10% in Canada. One writer rightly asks itself (Sunder, G., 2015) whether the facts described above show that

¹ ACL Service Ltd., *Fraud Detection Using Data Analytics in Government Organizations, discussion paper, 2010*, http://www.acl.com/pdfs/DP_Fraud_detection_GOVERNMENT.pdf

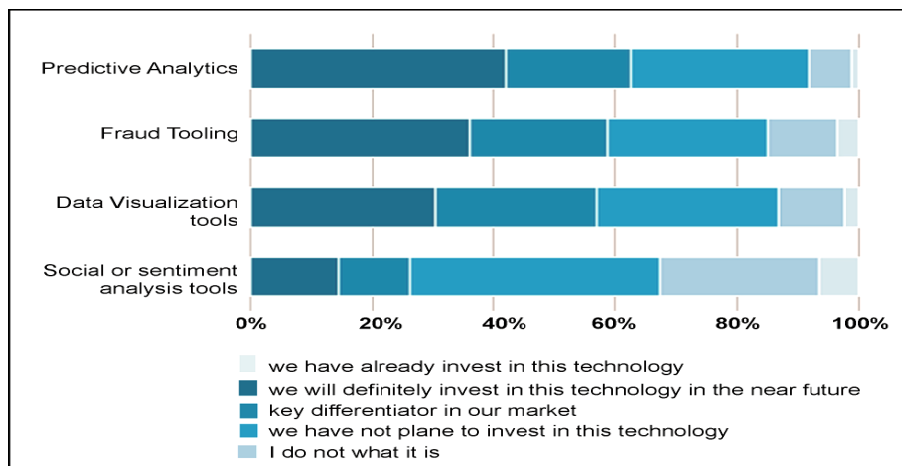
other methods of fraud detection (internal audits etc.) are more effective. From the first study result, correlated with the one of the Deloitte specialists from 2014, it can be shown that more than 26% of organizations use analytical tools for risk management. One can postulate that the low rate of detection may be associated with a large-scale underspending of data analysis processes in the prevention and detection of fraud.

Risk is the basic condition of insurance and is subject to any insurance contract. Risk is a danger, a threat to goods, people, businesses and for which the insurance companies can provide protection. Insurance risk arises as an essential compulsory element because: it is a future event, possible but uncertain, having as subject goods, capital, life or health; its initiation is not achieved by intentional means of the insured one; without this element there can be no valid insurance report.

Beside the risk deliberately assumed by the insurer, the risk of fraud is the highest, since it is a willful and premeditated act and often thoroughly prepared.

The main advantage in this case is to manage the risk of fraud. In our view, data analysis may help mitigate this risk by analyzing the data contained within the insurance files, performed in a certain period of time, in order to identify anomalies within their contents.

Figure 2 – The share of insurer's investments considering the types of fraud detection technologies (178 respondents)



Source: <http://www.celent.com/reports/market-dynamics-claims-fraud-detection>

In *The Market Dynamics of Claims Fraud Detection: Deal, Functionality, and Technology Trends*, Celent provides an overview of the market for claims fraud detection systems. The report describes the anticipated evolution of claims fraud detection system deals and looks at how vendors will try to find better solutions in terms of both functionality and technology.

The results obtained using classical methods of control are relative, leaving room for subjective interpretations.

Preparations for implementing and/or developing data analysis capabilities must be preceded by a detailed assessment of the current needs of the company and establishing the adaptability of the insurers to the information society (I-society). From the latter perspective, economic operators may be in the following situations:¹

Where Does Your Institution Rank in Fraud Prevention?

- Level 1 companies aren't using any tools and can't measure the sources of fraud or the magnitude or direction of risk.
- Level 2 companies have processes and tools, but still base investigations on intuition, tend to keep information siloed, and may not properly apply or understand the technology outside the immediate project team. And the rich demographic data the marketing side uses to acquire and retain customers isn't used to help understand the customer from a risk perspective. Measuring success isn't common.
- Level 3 companies have defined processes and use tools to acquire data and assess fraud and risk. However, the *ad hoc* tools used to acquire data and manage fraud and risk may not be integrated with one another to make cross-channel communication and measurement effective.
- Level 4 companies have a managed and measured approach. These institutions can benchmark themselves against industry performance, have processes in place to understand and root out new types of fraud, and consider fraud potential prior to entering new businesses.
- Level 5 companies are optimized. A portfolio approach is used to aggregate enterprise-level, cross-channel payment risk. These companies are completely up to date with regulatory compliance and relevant regulatory guidance.

¹ http://www.sas.com/news/intelligence_quarterly/q312.pdf

3. Techniques and tools for detecting and preventing fraud in auto insurance

The data and information analyses are influenced by the kinds of decisions that are taken inside an organization: **strategic** – *at the level of top management, their impact on the organization's resources and processes as well as the socio-economic context of the conduct of commercial activity. E.g.: changes in legislation, investment in certain resources, professional training, amendments to the internal working procedures, changes in information collection, new market entries etc.)* and **operational** – *management executive-level, and their impact on work processes at the executive level). E.g.: decision to refer suspicion of fraud to the competent authorities, assessing referrals/complaints of fraud, compensation decisions etc.)*.

In order for each type of decision to be substantiated at different institutional levels, analysis of data and information provide information products. Information products are the result of methodological approach of information analyst (*embodied in the evaluation report and analysis report*) that marks the completion of the analytical approach to the information within the **intelligence-led management** (the management based on information – requires a cycle with the following steps: analysis, decision, action (implementation of the decision), evaluation and monitoring. This process is not an analytical approach, but it is a way of managing business, a company culture that focuses on the data and information as key resources).

Information products are: strategic and tactical assessments, target profiles and problem profiles. Obtaining information products is carried out by qualified personnel (occupational field – **intelligence analyst** – COR group 2422, "specialists in the field of administrative policies"), who apply the standards of information analysis in the process analysis.

The analytical product should not be mistaken for the information product, although similar in concept and very often confused, in the literature as well. Analytical product belongs to "the analyst world" and information product is "manager's world". The connection between the two products is that the analytical products are part of the information product. An intelligence analyst may choose to substantiate a decision using multiple analysis techniques and then obtain a series of analytical products further used for generating information product).

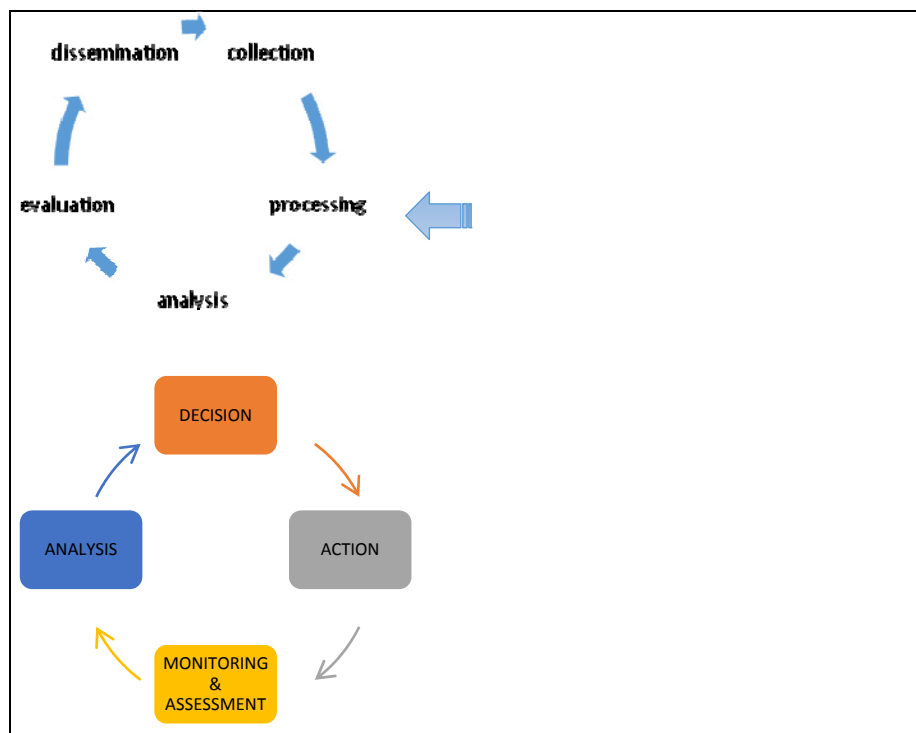
Table 1. Information products in relation to the analysis subject and to the institutional level of decision

Analysis subject \ institutional level	Strategic	Operational
Targets (individuals and groups)	Strategic Assessments	Targets profiles
Phenomenon	Tactical Assessment	Problem profiles

Source: Ratcliffe, J. (2009), "Strategic Thinking in Criminal Intelligence", 2nd Edition, Federation Press, p. 60.

The information analyst has several techniques and tools for data analysis and, depending on the type and subject of analysis, he may follow a certain scientific-practical approach for obtaining information products. The outcome of applying an analysis technique is the analytical product.

Figure 3. The relationship between the management process and analysis process



Source: authors' interpretation of above – mentioned concepts.

The application of analytical technique involves the use of analytical tools and processes: predictive and risk analysis (*based on historical data*), behavioral patterns (*used to detect changes in normal models*), advanced methods and link method (*the method determines whether there are network connections between diverse entities*).

Starting with the graphical representation of the analytic process (Figure 3), the system to be implemented should facilitate the following activities (competence elements of the occupational standard of information analyst – COR 2422 were also considered):

- planning activities related to the analyzing process:
 - ✓ defining the problem to be solved, the investigational area which shall be subject to analysis (*although it seems to be a logical step, we point out that many times the effects of fraud are not always visible*);
 - ✓ acceptance of the analysis theme, following the discussions with the applicant and/or recipient of the analysis;
 - ✓ a negotiation of the reference terms ;
 - ✓ development of a data collection plan (*identifying and determining the sources of necessary data for the analysis: public records (judgments, records of mortgages, tax/criminal records), telematics, Internet, open sources, social environments, talks with customers, public infrastructure, internal sources etc.*);
 - ✓ establishing the methodology of investigation, of methods, techniques and tools of work;
 - ✓ the establishment of the team members and support staff.
- create working assumptions (*it is a very important stage, and selecting the correct working paths will allow a moderate consumption of resources, time and energy*);
- carrying out the activities of the proper data analysis (using the methods and techniques of analysis):
 - ✓ formatting data, processing for analysis;
 - ✓ a homogenization of data of the same type, classification and clusterization;
 - ✓ verification of data sets from databases in terms of their consistency and completeness. Empty or unnatural records negatively influence the whole process of analysis and, by default, the analytical finished product;
 - ✓ data import in the analysis software;

- ✓ set of specific attributes for each entity, and the links between them;
- ✓ running various tests to look for symptoms, signs of fraud;
- ✓ running of multiple sequences to identify issues of interest;
- ✓ a preliminary data will be compared with other bits of data;
- ✓ a data comparison process plays an important role in analyzing the data;
- ✓ identify, as appropriate, of a representative set of comparable items. The usefulness of this measure is to use positive and negative experience gained before by teams of analysts;
- ✓ a re-evaluation of the results;
- ✓ an export of this data in a specific format;
- development of an analytical product (creating the fraud profile (will be taken into consideration both the person and the current operating mode));
- the assessment of the analytical product;
- decide how to use the analytic product (beneficiaries' proposals);
- disseminating the analytical product for the beneficiaries
- results, operating modes, entities involved in illicit activities may be archived and used later in the analysis, as elements of comparison and positive practice.

4. Choice of industry-specific analytic solution for the auto insurance field

Choosing a software application is an approach that should be taken into account the specific scope and expectations of management. A rigorous documentary research is recommended, to identify the "state of the art" of the technical solutions. Probationary periods offered by manufacturers are extremely useful for the purpose of running simulations and tests.

Although there are not solutions covering the entire range of situations generated by forms of fraud, any minimal (basic) package, for the construction of a system for the prevention and detection of fraud should include: the conversion software; a suite like Microsoft Office or equivalent; add-ins for the suite, allowing the increasing capabilities of analytic utilities; other cleaning, combination or processing of raw data tools. The most important tool is data analysis software. It is recommended to use a data analysis software to allow for the importation of various types of files (e.g., dBASE, Microsoft Access, Microsoft Excel, Adobe PDF, ODBC, text delimited, XML). The import interface must be friendly, contain

elements of selection/customization of importation, must be as intuitive as possible, in order to facilitate the activity in question.

There is not a toolkit to start a business fraud detection, it is not recommended to spend too much time selecting the perfect option. Just get started fighting fraud, use free and payable software, a combination of statistical, data visualization, data mining, and filtering tools. Although the software are not cheap, there is the possibility to maximize the benefits offered by the Office package (Excel, Access) or ActiveData for Excel (Bănărescu, A., 2014).

PowerPivot for Excel is a new technology aimed at providing self-service Business Intelligence (BI). PowerPivot is a real revolution inside the world of data analysis because it gives you all the power you need to perform complex analysis of data without requiring the intervention of BI technicians. This tool, an Excel add-in, implements a powerful in-memory database that can organize data, detect interesting relationships, and give you a swift way to browse information.

These are some of the most interesting features of PowerPivot: the ability to organize tables for the PivotTable tool in a relational way, freeing the analyst from the need to import data as Excel worksheets before analyzing the data; the availability of a fast, space-saving columnar database that can handle huge amounts of data without the limitations of Excel worksheets; DAX, a powerful programming language that defies complex expressions on top of the relational database. DAX allows you to define surprisingly rich expressions, compared to those that are standard in Excel; the ability to integrate different sources and almost any kind of data, such as information from databases, Excel worksheets, and sources available on the Internet; amazingly fast in-memory processing of complex queries over the whole database (Russo, M., Ferrari, A., 2011).

The investigation that I conducted on analytical products market revealed that manufacturers tend to offer high-performance products. One of the main advantages that a data analysis software may confer is its accessibility including non-users/experts in programming or economic statistics. At the same time, for more experienced analysts there are advanced features included for data analysis that involve tools with a high degree of sophistication.

Professional or dedicated data analysis software, such as ACL, Arbutus and IDEA are specifically designed for use with large and very large data sets. Features of this type of software: the data source is protected; can provide quick analysis; retains audit trails; built-in data analytical functions; user friendly; can import from various data sources and file formats; able to analyze 100 percent of transactions; field statistics; various types of sampling techniques; Benford's law

analysis; correlation and trend analysis; drill-down features aging; stratification; fuzzy matching; sophisticated duplicate testing; auto-run or automated procedures (Sunder, G., 2015).

Assuming that the data obtained have a format agreed upon in terms of quality (*accuracy and relevance*), the analytical techniques can be effectively implemented, I2 Analyst's Notebook being one of the working tools.

An effective method for identifying fraud with this analysis tool consists in investigating historical data stored in the databases of the insurer. One can assume that the claimants for undue compensation premeditate their actions, and often they become repetitive. By running in the data analysis, the identification of possible connections and fraud behavior (patterns) are envisaged.

If we take into account the fact that the working algorithms run a large volume of analytical units (thousands or even millions), the main advantage is that we need less time to identify fraud. In addition, the application has an interface (wizard) that guides users step by step to import the data and subsequently in the course of the analysis itself.

The data can be taken from a multitude of sources (spreadsheets, data connection, text files, etc.) as well as from other information systems through the use of open standards. The data can be rearranged and organized in analytical software, but we recommend cleaning and systematization of data by a tool like Microsoft Excel (the application provides data processing facilities: synthesis, summarizing, sorting, filtering, validation, a wide variety of functions, macros etc.).

The trigger of the analysis may be a determined or generic one, as a result of the specific activity for delimitation of certain areas of risk or to identification of risk indicators.

One of the most important steps is to import data into the analysis tool. The first step is choosing a primary key among the unique identifiers of the relationship. The unique identifier is a collection of one or more attributes that identify, uniquely, each instance of a relationship. In many cases there is only one attribute. When no attribute can be found as unique identifier, you can concatenate multiple attributes. Unique identifiers meaningful in the real world are called natural identifiers (name, first name, date of birth etc.) and others (PNC, chassis series, registration number, etc.) surrogate or artificial (Oppel, A, 2010).

Steps you must follow (*generically referred to as import specification*), contain the necessary settings for data in a text file or spreadsheet to be modeled in specific utility analytical units. There are several ways in which the import can be performed, i.e. choosing a predefined specification, modifying a pre-existing specification or building a new one, as shown in the figure below.

Figure 4. Types of import specifications

Edit Import Specification - [Clipboard, Modified]

Define Columns Select Rows Column Actions Select Design Assign Columns Import Details

Use this page to select an import design that represents the relationships of the items you want to import.

Row	Marca	Tip	Serie caroserie
2	MERCEDES	S350	WDD2210661A0...
3	DACIA	1310	UU1R11700H21...
4	RENAULT	CLIO	VF1UB0BCS3377...
5	MERCEDES	GLK	WDC2049841F7...
6	SKODA		TMB0C23UB890...
7	HYUNDAI	ACCENT	NLHCHM41AP720...
8	JEEP	GRAND CHEROKE	1J8HCE8M66Y1...

Network of telephone calls

The number of calls occurring in each direction between telephones.

Sequence of telephone calls

Telephone calls in chronological order.

Network of transactions

The number of transactions occurring in each direction between accounts.

Sequence of transactions

Account transactions in chronological order.

Sequence of events

A timeline with each event frame in chronological order.

More complex association chart

Relationships between three or more entities.

Timeline

A timeline with each link in chronological order.

Blank design

Create your own chart structure.

Source: I2 Analyst's Notebook screenshot

Broadly speaking, the import specifications contain a range of information on: the location where the data will be retrieved, the structure of the data to be imported, the types of entities and relations between them, their properties and attributes, time settings. For a correctly established import (*without any error of*

construction), the specification of import settings should be adjusted, after a preview of entities and links. With a view to saving time for data import, it is recommended that you save the import specification that may be used in importing the same files updated or new data, in a similar format.

Figure 5. The design of an import specification

The screenshot displays the 'Edit Import Specification' window. At the top, there is a toolbar with options: Select Worksheet, Select Rows, Column Actions, Select Design, Assign Columns, Date & Time Format, and Import Details. Below the toolbar is a table with the following data:

Row	Nume Pren...	CNP	Cuantumuldesp...	Locul prod...	Marcă	Seria mașinii	Nr. înmatriculare	Dosar daune	Data nașterii
2	icea Anton	1780412678789	30000	București sector 1	MERCEDES S350	WDD2210561A0...	B-02-NEF	00136k	11.12.1992
3	icea Anton	1780412678789	35000	București sector 1	MERCEDES GLK	WDC2049841F7...	B-03-SRD	00137a	11.12.1992
4	icea Anton	1780412678789	25000	București sector 1	MERCEDES GLK	WDC2049841F7...	B-03-SRD	00137a	11.12.1992

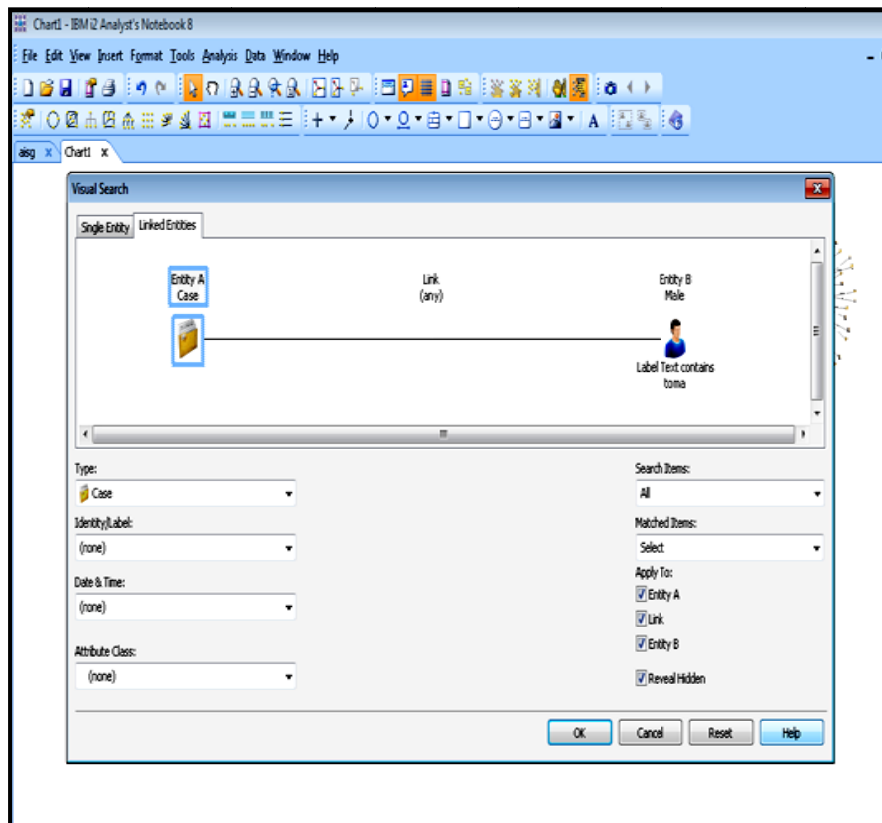
Below the table, there is a diagram showing the relationships between entities. The entities are represented by icons and labels: [Telefon], utilizator, [Nume Prenume], [Locus producerii accidentului], [Dosar daune], and [Dossier asigurate]. The relationships are indicated by arrows: [Telefon] points to utilizator; utilizator points to [Nume Prenume]; [Nume Prenume] points to [Locus producerii accidentului] (labeled 'accident'); [Nume Prenume] points to [Dossier asigurate] (labeled 'dosar asigurate'); [Locus producerii accidentului] points to [Dosar daune]; and [Dossier asigurate] points to [Dosar daune].

Source: I2 Analyst's Notebook screenshot

After completing these steps and the import is complete, one can use different procedures and analytical methods (full search text – is one of the simplest form of analysis, which allows for words or/end phrases identification).Running a

custom query, which contains a representation of the terms which we would like to identify, leads to identification of relations between analytical units.

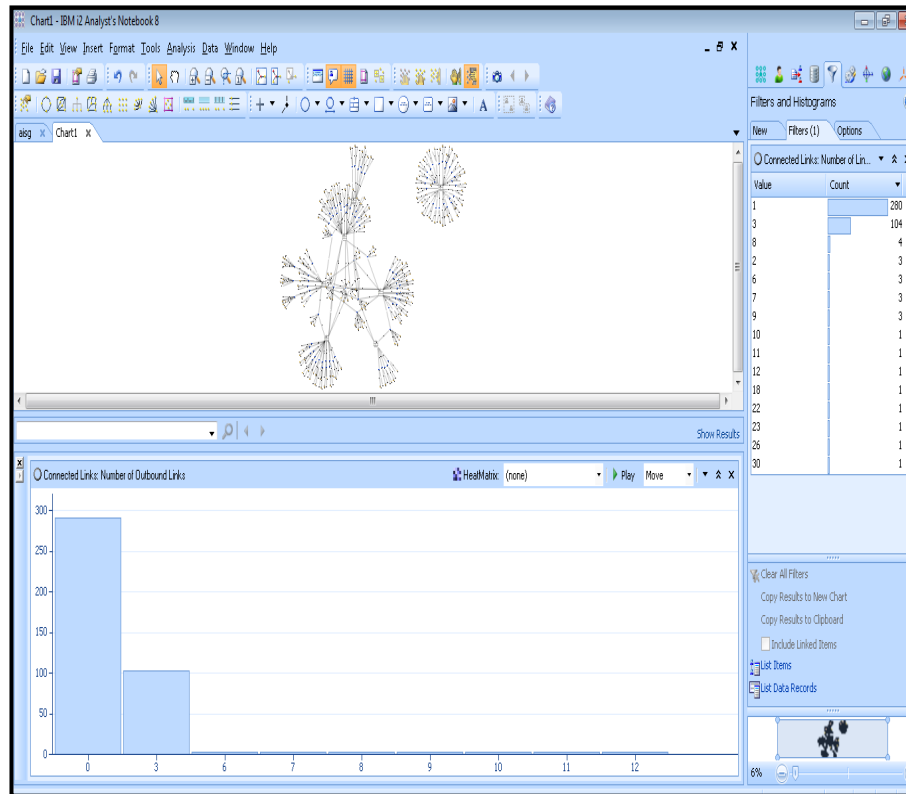
Figure 6. Example of query specification



Source: I2 Analyst's Notebook screenshot.

The imported data may be submitted to a series of filters so that the amount of data should be reduced to relevant samples (*only the elements which meet certain criteria will be shown*). The data type will be displayed in a histogram (a diagram showing the columns that display frequency data) by using the Histogram tool of the programme.

Figure 7. Example of data filter

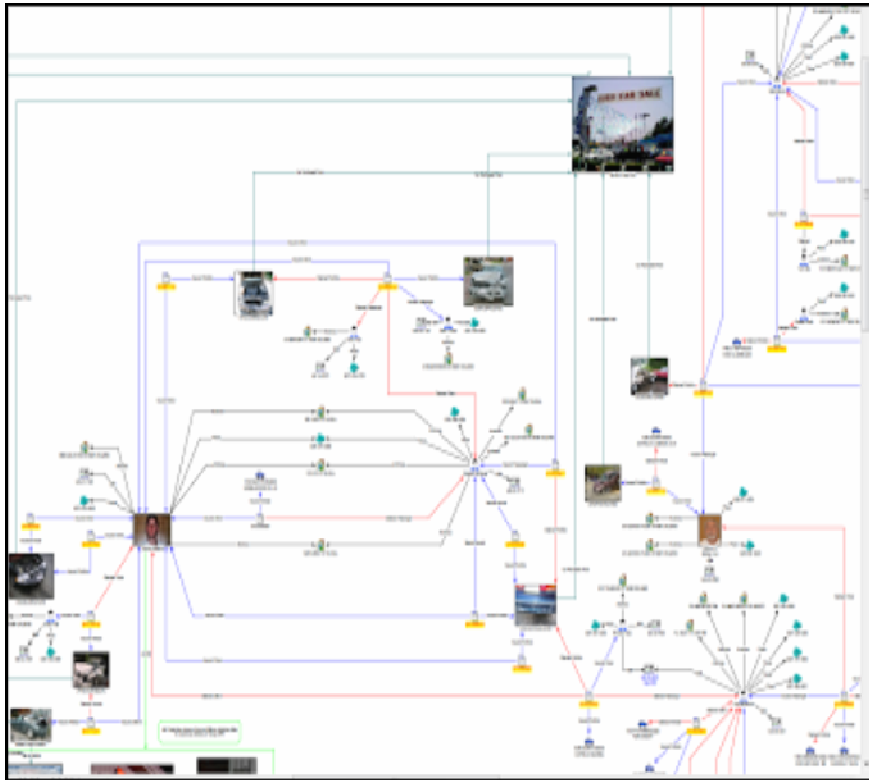


Source: I2 Analyst's Notebook screenshot.

The multiple procedures of queries results display enables the analyst to understand the procedures used by people who commit fraud, and the identification of fraud networks. Moreover, the use of several methods of representation of information in a chart, contributes to the development of dynamic thought processes.

In the end, the result of analysis of the links will be materialized in a relational map, like the one shown below.

Figure 8. Relational map displaying common links between members of an organized group



Source: <http://linkanalysisnow.com/2010/01/using-geospatial-analysis-to.html>

Other tools might be GeomediaProfessional and Incident Analyst. These spatial analysis tools provide the Analyst with a series of tools for capturing, processing, transforming, analyzing and visualizing geospatial information. The form of representation used is the map.

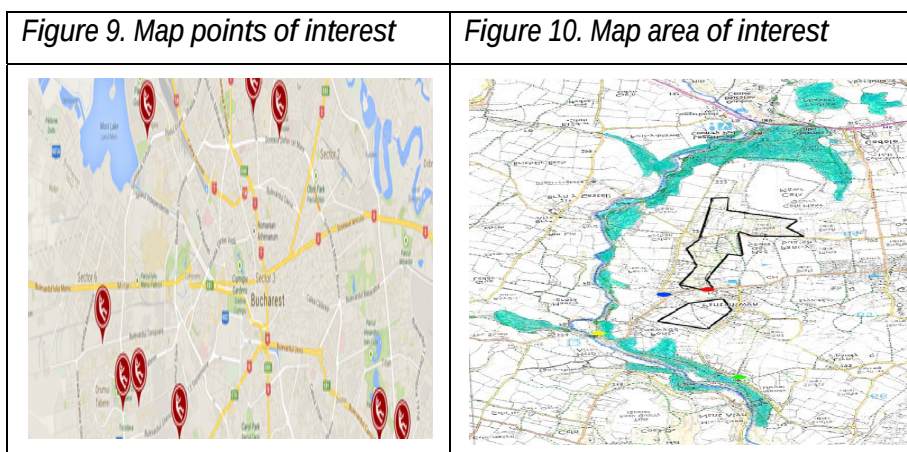
Geomedia suite offers a range of facilities for data and information analysis such as: connectors to various databases, functionality of connection data sets based on unique keys (join), data transformation and production of derived data, filtering and operating functionalities as well as visualization tools.

The key element in the geospatial analysis remains the location. Any information that contains data with geospatial content can be represented on the map and they can operate with specific spatial algorithms (including proximity, relationship etc.).

Geomedia suite offers the following types of methods for presenting information with spatial content descriptive maps, thematic maps, hotspots, timely transformation, real-time monitoring.

The table below lists a number of examples of how the information related procedures are applied, not necessarily made with Geomedia suite.

Descriptive maps – the simplest form of symbolic representation of reality recorded with data at their disposal. Simple representation of information as a map can bring added knowledge by identifying spatial relations between entities in a simple manner.



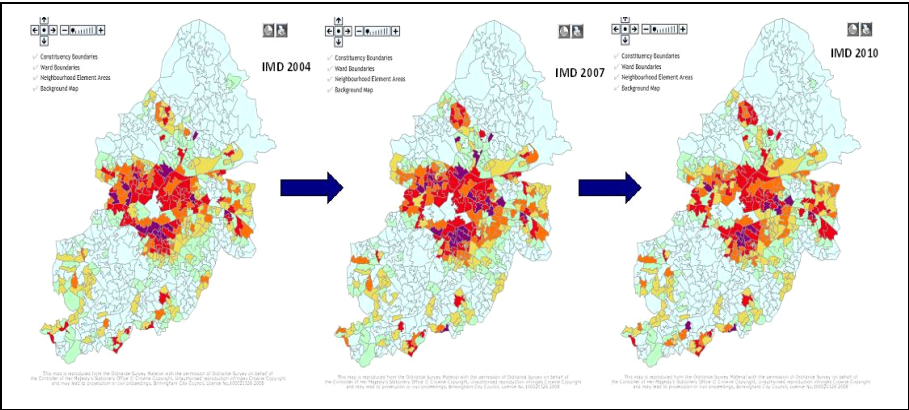
Source: <http://www.glennmorris.org.uk/map%20of%20alyn%20valley%20woods.htm>

<http://www.automarket.ro/stiri/harta-trecerilor-de-pietoni-periculoase-din-bucuresti-cartierul-drumul-54055.html>.

Thematic maps-context space representation of information specific to (some) variables in the analysis. Depending on the variable type (discrete or continuous) thematic maps can be made to view spatial information.

Depending on the tools animation images can be used, so that the analyst can identify the dynamics of the phenomenon. The advantage of this manner of visualizing the contents consists in knowledge improvement by introducing a temporal dimension in conjunction with a spatial one.

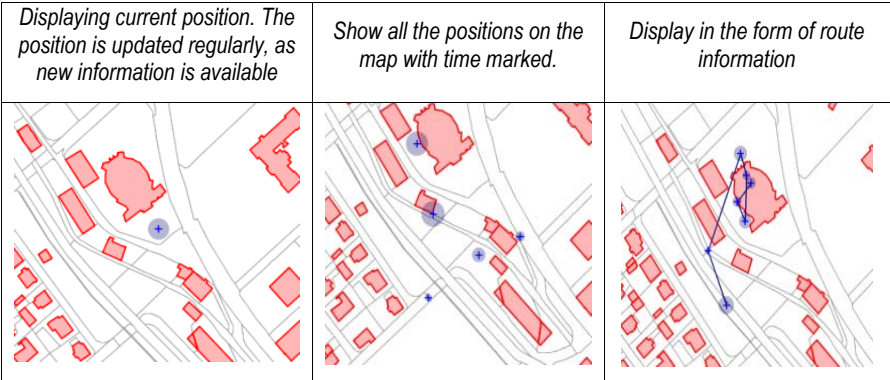
Figure 14. Example of dynamics map of the phenomenon



Source: <http://wiki.bebirmingham.org.uk/index.php?title=Deprivation>, last visit 20.04.2015

GPS-a Visual representation of the route mileage of a monitored target. This form of Viewer is important for analysis aimed at understanding the target itinerary and the context in which it was established.

Figure 15. Example of target itinerary



Source: [http://smartclient.intergraph.at/documentation/GPS_\(Microsoft_Location_API\)](http://smartclient.intergraph.at/documentation/GPS_(Microsoft_Location_API)), last visit 20.04.2015

Conclusions

Information resources of insurance companies should be harnessed by implementing advanced models of data and information analysis, in order to generate information products of strategic and operational level, to prevent and detect internal and external fraud.

Therefore, the auto insurance industry should make a sophisticated utility of data exchange and analysis fully operational, which could streamline tracking those groups specialized in obtaining money illegally, by invoking certain damage caused intentionally.

Simplification of the procedures for data communication will allow for an operational decision taken in due time. Such a utility would facilitate relational analyses with respect to environmental entities internally and/or externally, and, based on theme, a person or group of persons would be placed in a category of risk of fraud. By analyzing the attributes of these entities elements lead to proof of fraudulent activities can be identified.

Using visual and geospatial analysis facilitates the understanding of the relationship between analytical units as well as the understanding of the relevance of the location of the events. The ultimate goal of the cumulative use of the two types of analysis is to identify the fraudulent behavior of auto insurance.

After running several algorithms and batteries of tests, analysts can make a number of work assumptions such as: alleged author was not on the scene of the accident or at the time stated in the damage report; the person or group of persons have been involved in several accidents, in a relatively short period of time; the people involved live/reside in the same area with the possibility to know each other very well; a car is subject to several cases of damage (multiple reporting). At the same time, key people, structure and hierarchy of a network of fraud, events of interest, a series of connections and patterns of fraud will be identified.

Using social network analysis in combatting fraud activities contributes to the discovery and prevention of fraud crimes which have become more and more sophisticated, designed at high resolutions. Investigators can detect fraud patterns, exceeding the traditional limitations of an individual approach in a case of fraud, for a given transaction.

Data collection would be facilitated by the creation of an online portal for completing identification data and other specific elements, necessary to build up a damage claim case. This data might later be corroborated with other data from

internal or external sources (e.g. the known incidents of fraud). By applying risk-weighted aggregates and probability analysis specific situations are identified, which will be exploited in the course of internal investigations.

The immediate result of these proceedings would be a firm analytical reaction, followed by alerts based on existing black-lists, anomalies detection, monitoring, reporting of certain risk factors, avoiding paper format, processes automation which normally are performed manually, removing errors from further processing, automated supply of database.

Also, the lack of staff with tasks of verification and control can be replaced with by increasing the analytical capacity of fraud detection. In case of large amounts of data either multi-disciplinary working groups can be established, or the working group capacity may be combined with facilities offered by analytical solutions that can be easily integrated into existing information systems.

Creating a global framework for functioning and management of fraud should be put into effect on many levels, and in time, in order of priority and immediate needs of insurance companies.

Acknowledgement

This paper has been financially supported within the project entitled "SOCERT. Knowledge society, dynamism through research", contract number POSDRU/159/1.5/S/132406. This project is co-financed by European Social Fund through Sectoral Operational Programme for Human Resources Development 2007-2013. Investing in people!

Bibliography

- Alexandru, D. I.(2010), "Manual de Inteligența Afacerilor", publisher: lulu.com , ISBN-13: 978-1446694251.
- Bănărescu, A. (2014), "Detecting and Preventing Fraud with Data Analytics", *Emerging Markets Queries in Finance and Business – EMQFB*, București, 2014.
- Bolton, R., Hand, D. (2002), "Statistical Fraud Detection: A Review (With Discussion)", *Statistical Science*, 17(3): 235–255.
- Burge, P. &Shawe-Taylor, J. (2001), "An Unsupervised Neural, Network Approach to Profiling the Behaviour of Mobile Phone, Users for Use in Fraud Detection", *Journal of Parallel and Distributed Computing*, 61: 915–925.
- Cofan S.M., Ivan, L., et. al. (2014), "Analiza Informațiilor", Manual, ed. Ministerului Afacerilor Interne, ISBN 978-973-745-129-3.

- Cox, K., Eick, S. & Wills, G. (1997), "Visual Data Mining: Recognising Telephone Calling Fraud. Data Mining and Knowledge Discovery", 1: 225–231.
- Fawcett, T. (1997), "AI Approaches to Fraud Detection and Risk Management", Papers from the 1997 AAAI Workshop. Technical Report WS-97-07. AAAI Press.
- Palshikar, G.K. (2002), "The Hidden Truth – Frauds and Their Control: A Critical Application for Business Intelligence", *Intelligent Enterprise*, vol. 5, no. 9, pp. 46–51.
- Green, B. & Choi, J. (1997), "Assessing the Risk of Management Fraud through Neural Network Technology", *Auditing* 16(1): 14–28.
- Mantone, P. S. (2013), "Using Analytics to Detect Possible Fraud: Tools and Techniques", ISBN-13: 978-1118585627.
- Michalski, R. S., I. Bratko, and M. Kubat (1998), "Machine Learning and Data Mining – Methods and Applications", John Wiley & Sons Ltd.
- Miron C., (2012), „Infrațiunile contra patrimoniului prin nesocotirea în crederii”, Teză de doctorat, Universitatea „Nicolae Titulescu” , Facultatea de drept, Departamentul de științe penale și drept public, București.
- Murad, U., Pinkas, G. (1999), "Unsupervised Profiling for Identifying Superimposed Fraud", *Proceedings of PKDD'99*.
- Nigrini, M., (June 2011). "Forensic Analytics: Methods and Techniques for Forensic Accounting Investigations", Hoboken, NJ: John Wiley & Sons Inc. ISBN 978-0-470-89046-2.
- Nițu, I., (2012), "Analiza de intelligence. O abordare din perspectivă teoretică și aplicativă", Editura Rao, ISBN: 978-606-609-373-6.
- Oppel, A. (2011), "Modelarea datelor. Ghidul începătorului", Editura: Rosetti Educațional ISBN 978-973-7881-63-2.
- Phua, C., Lee, V., Smith-Miles, K. and Gayler, R. (2005), "A Comprehensive Survey of Data Mining-based Fraud Detection Research", Clayton School of Information Technology, Monash University, <http://arxiv.org/ftp/arxiv/papers/1009/1009.6119.pdf>.
- Ratcliffe, J. (2009), "Strategic Thinking in Criminal Intelligence", 2nd Edition, Federation Press, p. 60.
- Russo, M., Ferrari, A., (2011), "PowerPivot for Excel 2010, Give your data meaning", Microsoft Press, Redmond, Washington, 2011.
- Spann, D. (2014), "Fraud Analytics: Strategies and Methods for Detection and Prevention", John Wiley & Sons, ISBN-13: 978-1118230688.
- Sunder, G. (2014), "Fraud and fraud detection: a data analytics approach", (Wiley corporate F&A series), ISBN 978-1-118-77965-1 (hardback) — ISBN 978-1-118-77967-5 (ePDF) — ISBN 978-1-118-77966-8 (ePub) — ISBN 978-1-118-93676-4 (oBook), Printed in the United States of America.

Internet sources:

- <http://www.dailybusiness.ro/stiri-finante-banci/ce-vor-face-asiguratorii-pentru-a-nu-mai-fi-fraudati-pe-segmentul-casco-100120/>
- <http://www.accenture.com/SiteCollectionDocuments/PDF/Accenture-How-Effectively-Fight-Insurance-Fraud.pdf>
- http://www.sas.com/news/intelligence_quarterly/q312.pdf
- <http://linkanalysisnow.com/2010/01/using-geospatial-analysis-to.html>

www.hexagongeospatial.com

<http://wiki.bebirmingham.org.uk/index.php?title=Deprivation>

[http://smartclient.intergraph.at/documentation/GPS_\(Microsoft_Location_API\)](http://smartclient.intergraph.at/documentation/GPS_(Microsoft_Location_API))

PricewaterhouseCoopers LLP (2011). "2009 Global Economic Crime Survey", https://www.pwc.com/en_GX/gx/economic-crime-survey/pdf/global-economic-crime-survey-2009.pdf