



**Silvia Amatucci,**  
Professional Actuary of the Statistics  
and Actuarial Accounting  
Department of the National Institute for Insurance against  
Accidents at Work (INAIL, Italy)

## ACCIDENTS AT WORK: A SEMANTIC ENGINE TO CLASSIFY THEM

ESAW (European Statistics on Accidents at Work) is the system for coding accidents at work, born in Europe in the early 1990s and adopted in Italy since 2001.

ESAW compares the cause of the accident phenomenon between different Member States of the European Union, standardizing the language with which the causes and circumstances of accidents are described. It also has the purpose of evaluating and identifying interventions to improve the conditions of health and safety of workers in compliance with EU directives up to reducing the frequency and severity of the injuries themselves. To counter an event, it is necessary to know how and why it occurs. The icon aims, therefore, at the constant improvement of data quality: the more reliable the information, the better the prevention policy adopted will be. To this end, INAIL has exploited the potential of the semantic engine IRIDE is able to reduce the uncertainties due to operators in the coding phase of the accident report, and to standardize the process. The innovative approach, the process and the results: the ESAW-IRIDE database The innovative approach of the IRIDE software is in the support of the coders in the attribution of the correct code to the variables that describe the causes and the circumstances of the accident at work. The basic principle of semantic technology is the interpretation of the natural language used in the accident report. In other words, these engines are able to include the contents of the texts and then assign, through the so-called «rules» defined by experts, a code that identifies a word or a concept. Initially, the accident reports for specific sectors of economic activity were analyzed, in addition to defining the domain on which IRIDE bases its rules. In 2018 the project successfully leads to important results in terms of coding quality and finally today we can talk about the IRIDE-ESAW database.

### Introduction

Among the strategic objectives of INAIL (the Italian insurance institute against accidents at work), for many years now, there has been the Protection against accidents at work, which provides interventions for safety and prevention.

In this mission, which is fundamental for the improvement of the Institute's information system, the contribution of the IRIDE-ESAW<sup>1</sup> (IRIDE from now on) project was significant, which envisaged how the new classification/coding of the ESAW/3<sup>2</sup> variables was supported through a facilitator system, in this case a semantic engine, capable of improving the quality of the recorded data, simplifying the process of such registration and reducing time and economic costs deriving from the processing of a claim.

In 2016 this project «The semantic engine ESAW-IRIDE: Accident analysis in support of prevention» received the ISSA Certificate of Merit of Good Practices at an international level.

The IRIDE project has therefore placed itself at the forefront at national and European level as the first experience applied through the use of so-called artificial intelligence.

### Research methods

ESAW (European Statistics on Accidents at Work) is the system for coding accidents at work, born in Europe in the early 1990s and adopted in Italy since 2001.

ESAW compares the cause of the accident phenomenon between different Member States of the European Union, standardizing the language with which the causes and circumstances of accidents are described. It also

<sup>1</sup> IRIDE-ESAW is a semantic engine to code accidents at work

<sup>2</sup> ESAW – European Statistics on Accidents at Work. ESAW/3 are variables that describe causes and circumstances of accidents at work

has the purpose of evaluating and identifying interventions to improve the conditions of health and safety of workers in compliance with EU directives up to reducing the frequency and severity of the injuries themselves. To counter an event, it is necessary to know how and why it occurs. The icon aims, therefore, at the constant improvement of data quality: the more reliable the information, the better the prevention policy adopted will be. To this end, INAIL has exploited the potential of the semantic engine IRIDE is able to reduce the uncertainties due to operators in the coding phase of the accident report, and to standardize the process. The innovative approach, the process and the results: the ESAW-IRIDE database The innovative approach of the IRIDE software is in the support of the coders in the attribution of the correct code to the variables that describe the causes and the circumstances of the accident at work. The basic principle of semantic technology is the interpretation of the natural language used in the accident report. In other words, these engines are able to include the contents of the texts and then assign, through the so-called «rules» defined by experts, a code that identifies a word or a concept. Initially, the accident reports for specific sectors of economic activity were analyzed, in addition to defining the domain on which IRIDE bases its rules. In 2018 the project successfully leads to important results in terms of coding quality and finally today we can talk about the IRIDE-ESAW database.

This is a new way of coding cause and circumstances of accidents at work.

Employer has to report accidents at work. To do that they describe where the accident took place, what the worker was doing, why and how the accident occurred. These data are called ESAW variables. Yearly the Member States of the European Union have to send this kind of data to Eurostat.

ESAW compares the data of the accident phenomenon between different European Countries, standardizing the language with which the causes and circumstances of accidents are described. It also has the purpose of evaluating and identifying policies to improve the conditions of health and safety of workers in compliance with EU directives up to reducing the frequency and severity of the injuries themselves.

INAIL aims, therefore, at the constant improvement of data quality because the more reliable the information, the better the prevention policy adopted will be.

Since 2017, INAIL (National Institute for Insurance against Accidents at Work) is using a semantic engine to assist the encoder in coding ESAW variables. From then to today Esaw data has definitely improved.

ESAW is the European coding system of injuries, adopted by Inail since 2001, allowing to compare accidents at work rates in different Member Countries. Knowledge of causes and circumstances of occurrence of accidents at work is essential for planning policies for prevention and protection aimed at reducing the frequency and severity, as well as enabling the plan of measures to improve working conditions. Inail, in order to improve their data quality and reduce some elements of uncertainty due to subjective interpretation of ESAW variables/3, in 2010 has undertaken the way of technological innovation. Inail has therefore developed the software-semantic engine IRIDE aimed to an assisted encoding, and proposed a new

model for accident reporting, structured to highlight more information ensuring proper allocation of variables. As an injury is often the result of a cascade of adverse events and contributing factors, it becomes important not only to quantify it in a timely manner, but also to evaluate it through a qualitative characterization of the event. The software ESAW-IRIDE is therefore a timely response to the need of making an accurate and standardized encoding.

The ESAW project, which involves several European Union Countries, aims to harmonize methodologies and criteria for registration of data on accidents at work to make them comparable in the different Countries. The last phase of the project, called ESAW/3, aims to standardize the description of the causes and circumstances of the accident at work through a sequence of eight variables, which are used to represent the actual dynamic accident. The analysis of the data relating to these variables, if properly coded, provides a vision qualitatively and quantitatively reliable for injuries; while this activity allows an interpretation of the data in terms of prevention, aimed at proposing effective corrective measures. In Italy Inail adopted this encoding system since January 2001, thereby enhancing its role preventional.

In fact, the implementation of the database with information on terms of occurrence of injuries is crucial to define and/or propose policies to reduce the frequency and/or severity of harmful events. However a proper codification of the variables ESAW/3 – The Working Environment (the workplace, work premises or general environment where the accident happened), The Working Process (the general activity or task being performed by the victim at the time of the accident), The Specific Physical Activity (to describe the victim's activity immediately before the accident), The Deviation (an abnormal event, such as totally or partially losing control of a machine or falling onto/off something), The Contact — Mode of injury (it describes how the victim was injured and how he or she came into contact with the object that caused the injury) and Material Agents involved (tool, object or instrument used by the victim when the accident happened) - can be complex for the encoder in relation to the articulation of the same variables and because these variables normally derive from the interpretation of unstructured description of the accident at work contained in its complaint.

Since an injury is often the result of a cascade of adverse events and contributing factors, it becomes important not only to quantify it in a timely manner, but also to evaluate it through a qualitative characterization of the event that only an accurate encoding can ensure. The software ESAW-IRIDE is therefore a timely

response to the need to make an accurate and standardized encoding, raising the level of analysis of the accident's dynamics .

The good practice has the following objectives:

- an accurate analysis of injury rates;
- a uniform and rational encoding of injuries at national level because INAIL is distributed nationwide;
- to support more targeted interventions for prevention.

The innovative approach was to use software mainly to support the encoders to attribute the correct codes to the variables that describe causes and circumstances of accidents at work.

The principle on which, in general, technologies of semantic ontology are based consists of interpreting natural language; in other words they are able to understand the content and managing such knowledge at a conceptual level (not only through key words), in a way similar to what people do.

Using mathematical functions, so-called «rules», it is possible to assign the value of a code to a set of concepts.

A semantic engine IRIDE scans the text, it extracts the concepts and relationships, it transforms the knowledge into codes; starting from the analysis of the information contained in the complaints of injuries (an unstructured text), IRIDE worked out a series of codes that are suggested for the encoding of the case.

The encoder can not use the suggestions provided by the software, choosing, if deemed more suitable, a code other than those proposed.

In the engine also other functions are enabled, visible and usable according to the degree of user profiling (encoder, validator, administrator):

- “search”, which allows you to extract complaints interesting responsive to various parameters, including a «string» in natural language;
- “prompter”, allowing you, directly entering the texts that describe the causes and circumstances of the event, to experience the suggestion of real-time encoding ESAW proposed from the engine to each of the variables under consideration;
- “search of validated complaints”, which allows access to the history of all complaints used for training of IRIDE;
- “monitoring”, which allows you to monitor the ability of classificatory IRIDE, leading in the choice of more complaints to be validated and will be used in the future to support the statistical studies on the dynamics of workplace accidents.

As a first step it was created the domain on the basis of scientific and technical documents relating to the classification system introduced by ESAW. Subsequently, the system was implemented using encodings of real complaints, trying gradually to cover all the variable codes ESAW3. Real complaints were investigated in a timely manner by the group of technical experts who ensured for each code compliance attributed to the coding rules imposed in Europe.

When a high level of correspondence was reached between the encoding suggested by the engine and that one identified by the experts, it proceeded with a test phase by extending the system to INAIL local operators.

Monthly, in order to refine the rules of case studies or special codes that are scarcely recurring or not covered by the set of complaints encoded in the seats, a sample is extracted for the validation phase: tested timely, such a sample is used for the subsequent training of the system.

IRIDE is a useful tool to monitor the dynamics of «emerging» accidents, as well as those less serious events, and to characterize the «deviations» that led to the damaging event, comparing them with trends and experiences in different European Countries. In proper implementation of a prevention intervention, data analysis is not necessary only in the initial stage of design and implementation, but also at a distance of time, in evaluating the effectiveness of the intervention done.

INAIL has invested considerable financial, technical and organizational resources.

To implement these objectives, under the supervision of Central Directorate for Prevention, synergies between different INAIL departments (Technical Advisory for Risk Assessment and Prevention, Statistical and Actuarial Consulting, Innovation Technological Advisory, Centrale Directorate for Digital Organization) were fundamental.

To enrich the data of additional concepts, synonyms and coding rules, key step to improve the interpretative skills of the software, data were regularly extracted in order to evaluate IRIDE performance.

An interdisciplinary group of professionals (actuaries, biologists, chemists, engineers) was constituted to carry out the following activities:

- creation of the domain on the basis of scientific and technical documents relating to the classification system introduced by ESAW;
- system implementation through encodings complaints certified by Eurostat and then encodings real complaints relating to all sectors of economic activity;
- preparation of set encodings «certified»;
- clarification of some rules for situations that can lead to difficulties of interpretation;

- timely analysis of coding to ensure compliance with the codes assigned to the coding rules imposed in Europe;
- advice on questions and problems posed by the developers of the software;
- training and information for expert coders;
- information activities process managers;
- monitoring sample encodings performed by expert coders;
- preparation of FAQ for the most common data cases.

A more accurate reconstruction and interpretation of injuries, such as that IRIDE software helps to get, integrated from other data held by INAIL, can identify more closely the real safety needs of each production sector.

It gives, ultimately, a knowledge base of size and quality significantly higher than the current one, with which to evaluate, design and implement preventive interventions aimed at reducing the severity and frequency of accidents. Furthermore, given the overall objectives of ESAW, the Italian experience of IRIDE lends itself to being shared with other participating States at the coding system, in order to achieve better comparability of data and highlight any differences in the nature and distribution of accidents.

The IRIDE semantic engine is therefore able to understand the unstructured information present in accident reports and develop a range of codes to be suggested to the operator for the coding of the case.

IRIDE belongs to the latest generation of information technologies that interpret language and are able to achieve content understanding and manage knowledge not only by keywords but on a conceptual level, in a similar way to what people do.

In practice, IRIDE recognizes the contents of a complaint and leads them back to the basic concepts – ontologies – on which is developed: place, worker, action, object.

The implementation process of the engine was divided into 4 phases:

1. learning
2. training
3. classification
4. validation

and involved a central working group and 40 operators from the regional offices who used the system to codify complaints actually received by Inail: this made it possible, through a punctual control of the results, to further improve the performance of the system. Iride was introduced in 2017.

IRIDE has not replaced the figure of the operator, which has remained central to the coding process, but it has been able to reduce work times and consequently economic costs and to reduce the subjective interpretation of the operator.

## Results

The IRIDE engine has been trained through a series of ontologies, taxonomies, rules and real accident reports, relating to all economic activities, and validated by a group of INAIL central experts.

The sample of validated complaints was used both to instruct the semantic engine and to test its goodness. The levels of coverage achieved were more than satisfactory: almost all the codes relating to the five main ESAW/3 variables were included in the engine.

In detail, the results show that the engine returns the correct code on average in 91% of cases for the five main ESAW/3 variables and in 89% of the cases for material agents.

Furthermore, considering the single accident report, Iride, nine times out of ten, returns the correct code of at least 6 of the 8 ESAW/3 variables.

By focusing on the position of the correct code in the range of suggestions that the engine returns, IRIDE identifies the correct code in the first three suggestions, respectively in 96% of cases for the five ESAW/3 variables and in 84% of cases for material agents.

The excellent results achieved by the semantic engine have made it possible to replace the current manual coding system with the new system based on the IRIDE engine.

## BIBLIOGRAPHY

Marco Albanese – La codifica ESAW diventa motorizzata. Dati INAIL 2016

Maria Rosaria Fizzano – IRIDE: un motore per la prevenzione. Dati INAIL 2016

Silvia Mochi – IRIDE ESAW: i risultati. Dati INAIL 2016

Silvia Amatucci - Accidents At Work: a semantic engine to classify them. ISSA Berlin 2020

EUROSTAT - European Statistics on Accidents at Work (ESAW) - Summary methodology