Avoidance of operational sampling errors in drinking water analysis

Ana Fernandes (ID)[a], Margarida Figueiredo (ID)[b], Jorge Ribeiro (ID)[c], José Neves (ID)[d,e] and Henrique Vicente (ID)[e,f,*]

[a] CBIOS, Escola de Ciências e Tecnologias da Saúde, Universidade Lusófona, Campo Grande 376, 1749-024 Lisboa, Portugal
[b] Departamento de Química, Escola de Ciências e Tecnologia, Centro de Investigação em Educação e Psicologia, Universidade de Évora, Rua Romão Ramalho, 59, 7000-671 Évora, Portugal
[c] Instituto Politécnico de Viana do Castelo, Rua da Escola Industrial e Comercial de Nun'Álvares, Viana do Castelo, Portugal
[d] Instituto Universitário de Ciências da Saúde, CESPU, Rua José António Vidal, 81, 4760-409 Famalicão, Portugal
[e] Centro Algoritmi, Universidade do Minho, Campus de Gualtar, Rua da Universidade, 4710-057 Braga, Portugal
[f] Departamento de Química, Escola de Ciências e Tecnologia, REQUIMTE/LAQV, Universidade de Évora, Rua Romão Ramalho, 59, 7000-671 Évora, Portugal
*Corresponding author. E-mail: hvicente@uevora.pt

(ID) AF, 0000-0002-1358-1372; MF, 0000-0002-8969-0380; JR, 0000-0003-1874-7340; JN, 0000-0002-8863-0351; HV, 0000-0001-8456-7773

ABSTRACT

The internal audits carried out in the first half of 2019 in water laboratories as part of quality accreditation in accordance with ISO/IEC 17025:2017 showed a high frequency of adverse events in connection with sampling. These faults can be a consequence of a wide range of causes, and in some cases, the information about them can be insufficient or unclear. Considering that sampling has a major influence on the quality of the analytical results provided by water laboratories, this work presents a system for reporting and learning adverse events. Its aim is to record nonconformities, errors, and adverse events, making possible automatic data analysis aiming to ensure continuous improvement in operational sampling. The system is based on the Eindhoven Classification Model and enables automatic data analysis and reporting to identify the main causes of failure. Logic programming is used to represent knowledge and support the reasoning mechanisms to model the universe of discourse in scenarios of incomplete, contradicting, or even unknown information. In addition to suggesting solutions to the problem, the system provides formal evidence of the solutions presented, which will help to continuously improve drinking water quality and promote public health.

Key words: drinking water, Eindhoven Classification Model, Knowledge Representation and Reasoning, logic programming, sampling errors, water quality

HIGHLIGHTS

- An adverse event reporting and learning system for water sampling is described.
- The Eindhoven Classification Model is extended and adapted for water sampling.
- Logic programming is used for knowledge representation.
- The proposed system can deal with insufficient or ambiguous information.
- The system allows to identify the relevant issues behind the errors that may occur.

INTRODUCTION

Securing water quality for human consumption through a public supply system is an essential element of health policy. By the end of the 19th century, an assessment and control of risks to human health from transmission of diseases caused by water consumption had been carried out empirically on the basis of the physical appearance of water (Bagchi 2013). Epidemiological studies conducted by John Snow showed a close association between the consumption of water with fecal contamination and a cholera outbreak in London (Snow 1855). The discovery of the existence of microorganisms by Louis Pasteur in 1863 and the isolation of *Vibrio cholera* bacillus in 1883 by Robert Koch formed the crucial scientific basis for the association of water use with public health and served as a starting point for the establishment of practices and protocols to control water quality (Geison 1995; Brock 1999). Until the middle of the 20th century, the quality of water for human consumption had been largely assessed based on its organoleptic characteristics, i.e., its colorless, tasteless, and odorless features (Eaton *et al.* 2017). However, this type of assessment does not guarantee the protection of public health

from pathogenic microorganisms and hazardous chemicals. Therefore, it is imperative to establish standards based on quantifiable parameters that define the properties of water for human consumption in terms of microbiological, physical, chemical, and radiological factors (ISO/IEC 17025 2017). Current methods of controlling water quality are time-consuming and complex, and involve the use of huge technical and financial resources (e.g., calibrated equipment, infrastructure, skilled technicians, and reagents). With all this complexity, a large number of errors can occur (Byleveld *et al.* 2008). A failure can be defined as failure of a planned action and use of an incorrect plan, and these can be in connection with products, processes, and systems (Hommerson *et al.* 2008). The best way to prevent similar errors from happening again is to report them, i.e., create experiential learning systems to identify their causes. Although this approach is widely used in the water sector, it is important to note that classification systems work better when they are limited to a specific field/stage/phase (e.g., pre-analytical, analytical, and post-analytical phases). Sampling is an important stage in a water quality control program as it can affect results if a sample is not representative of the water being controlled. On the one hand, sampling errors can arise if the sampling is not designed to answer specific questions (usually regulators specify the sampling approach and the sampling design to be adopted). On the other hand, sampling errors can also be related to operational issues. In this work, a system for reporting and learning errors is presented, with the aim of describing and preventing those that occur in operational sampling of water for human consumption. The operational sampling of water for human consumption involves the following four main steps:

- Collection;
- Conditioning;
- Preservation; and
- Transportation.

These are shown in Figure 1. Water samples are delivered to the laboratory for analysis, and the sampling technician is responsible for taking a valid and representative sample (Li & Migliaccio 2011).

Due to the increasing importance attached to the representativeness of sampling, greater emphasis has been placed on the adequate collection, conditioning, preservation, and transport of samples. Thus, it is advisable that laboratories help to plan a sampling program in partnership with their customers. This partnership is essential to ensure that the selection of samples is adequate in order to minimize errors and to correspond to the needs of the consumer (Nollet & Gelder 2014). At this point, it is important to reiterate that the main goal of water analysis is to aid the decision-making process.
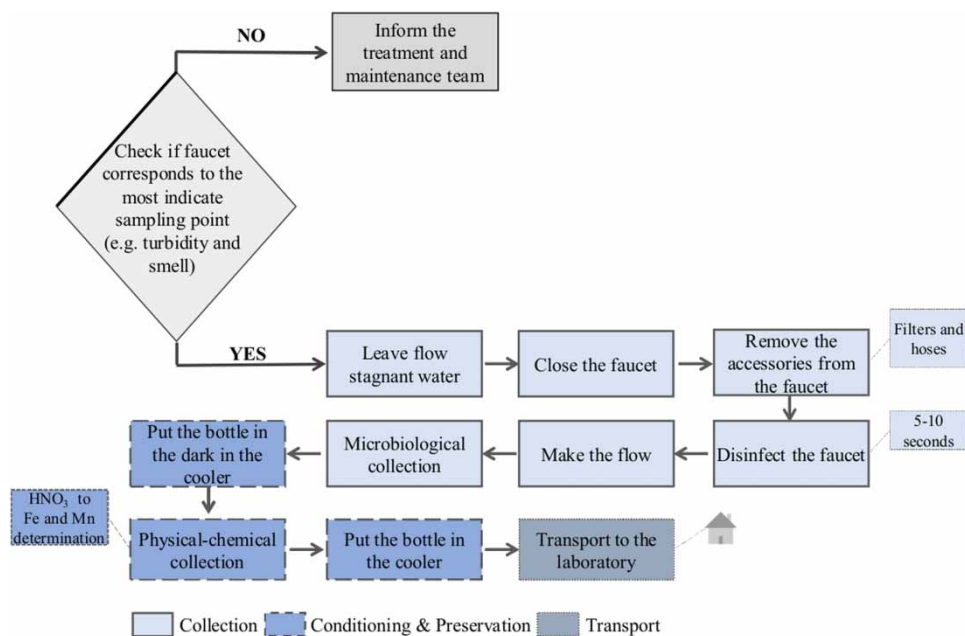
Figure 1 | The main steps in operational sampling of water for human consumption.

Thus, the reliability of the decisions depends on the variability of the experimental results. Despite variability being inherent to natural systems, if the variability due to errors in sampling, transport, handling, and/or analysis is underestimated, wrong decisions can occur with negative consequences for health, environment, and financial matters, just to name a few. Indeed, the errors that occur during the water analysis process are difficult to quantify. Inadequate training, reliance on automated systems, and complicated software resources are some of the relevant issues behind the errors that may occur (Reason 2000). Knowledge is crucial to solve the problems of the economy and society. In everyday situations, however, the information that flows from various sources is often incomplete, inaccurate, uncertain, contradictory, and even unknown. On the other hand, existing models are built based on some idealizations that remove these real-world properties.

The result is a system that, due to its inability to model the world or universe of discourse, never delivers the expected answers (Parsons 1996). In the water sector, particularly with regard to sampling errors, there are many situations where the information available on errors is insufficient or ambiguous. The system proposed in this work is based on the Eindhoven Classification Model (ECM) (van der Schaaf 1995) and focuses on avoiding the causes of errors by applying a model specially developed for the operational sampling phase.

The ECM was created to address human error in the chemical process industry and has since been applied in different areas (Henneman *et al.* 2006; Raab *et al.* 2006; Simmons & Graves 2008; Rodrigues *et al.* 2011; Vicente *et al.* 2015; Fernandes *et al.* 2019). The ECM contains two types of errors, namely, human or active errors and latent ones. With regard to human error, the ECM incorporates Rasmussen's SRK thematic, which includes three levels of behavior, namely, competence-based, rule-based, and knowledge-based levels (Rasmussen 1976). In the case of latent errors, the ECM differentiates between technical and organizational errors (van der Schaaf 1995). The former occurs when there is an obstacle associated with physical components (e.g., devices and physical installations), while the latter is associated with protocols, procedures, or knowledge transfers.

The main objective of this work is to present an adverse event reporting and learning system, developed specifically for the sampling phase. Based on the internal audits carried out in the first half of 2019 in water laboratories, the adverse event 'Failure in Sampling' was chosen to demonstrate how the proposed system can be applied to real-world problems. This system, known as Adverse Event Reporting and Learning System related to Water Sampling (AERLS-WS), intends to record the non-conformities, errors, and adverse events that occur during sampling. It also intends to be a learning system that enables data analysis to ensure continuous improvement in drinking water quality and the promotion of public health.

RELATED WORK

The need for accurate predictions of water quality has encouraged researchers to develop new methods and improve the predictive capabilities of traditional ones. On the other hand, the potential of Artificial Intelligence (AI)–based problem-solving methods has also been recognized significantly in the water industry. Several studies in the literature illustrate the application of these tools to quality water control, but very few relate to the drinking one. Indeed, water quality modeling is a powerful tool for studying, describing, and predicting the ecological status of an aquatic ecosystem. Morales-Torres *et al.* (2016) presented a study about the use of sustainable drainage systems to manage urban runoff, aiming to contribute to environmental and landscape improvement. The authors developed a tool to analyze the impact of storm water management on the urban environment and introduced energetic and environmental criteria in the decision-making process. Stefanovic *et al.* (2015) developed a web-based system to manage Serbian reservoirs. The authors created a software that brings together all the relevant actors, processes, data sources, and applications. The software allows to test different organizational scenarios, due to its flexibility, extensibility, and the fact that it could be customized. Allan *et al.* (2013) studied the effects of regulation, proprietorship, and business culture on risk management within the water sector. The authors stressed that a paradigm shift is necessary. The water sector should move from a reactive to a pro-active risk management culture based on knowledge. However, the authors stated that there is a gap in the understanding of the drivers required to bring about such a change. Pearson *et al.* (2010) presented a decision support system to support managers in the urban water sector. The proposed system aims to shift decision-making from the conventional standards to a sustainable, inclusive, and dynamic paradigm, anchored by social learning and commitment. The authors stressed on the basic units of social learning, i.e., knowledge transfer for transformation, monitoring and evaluation of the decision-making process, and stakeholder engagement.

In the water sector, there are many situations in which information is incomplete. However, most of the decision support systems ignore this fact. Most of the time such systems do not provide the expected answers due to their inability to represent

the universe of discourse. Thus, it is of utmost importance to deal with defective information. In this context, the use of the logic programming (LP) paradigm to express *theories* has grown in recent years, and many other fields have started using its tools and methodologies. *Theories* are typically expressed as a succession of rules and facts that allow to deduce non-logical consequences using logical inferences (Fernandes *et al.* 2020). According to the extension to the LP language presented in Neves (1984), both explicit (or strong) negation and explicit declarations may be used in the establishment of rules and facts. Furthermore, *theories* can be enhanced by adding special rules in the form of invariants to specify the context under which the universe of discourse should be understood. Such invariants specify the conditions that must be met and must have overall consistency, i.e., it must not be possible to infer a conclusion and its contradiction (Neves 1984; Kakas *et al.* 1998; Neves *et al.* 2021). This extension to the LP language uses the Open World Assumption (OWA), i.e., everything is unknown or undefined until one has some secure evidence of its truthfulness or falseness (Keet 2013). The OWA differs from the more usual Closed World Assumption, where everything is assumed false until there is solid evidence of its truthfulness (Keet 2013). The use of the OWA is the main advantage of the method used, since it allows to deal with incomplete, inconsistent, contradictory, default, or forbidden information, and the conclusions are an object of a formal proof. Indeed, this point is of utmost importance, since in the real-world, information is incomplete and updateable.

METHODS

The ECM was initially conceived to address errors in the chemical industry (van der Schaaf 1995) and has since been adapted to be used in different areas, e.g., steel industry, power generation, healthcare, water distribution, or public administration (Henneman *et al.* 2006; Raab *et al.* 2006; Simmons & Graves 2008; Rodrigues *et al.* 2011; Vicente *et al.* 2015; Fernandes *et al.* 2019). In this work, an adjusted version of the ECM was conceived, with extensions and adaptations for water sampling scenarios, as well as Causal Trees (CTs) that describe the root causes of errors. The data used in this work were obtained during the 1st semester of 2019, in internal audits carried out at water laboratories within the scope of ISO/IEC 17025:2017 quality accreditation. Checklists were used to record nonconformities. In the checklists, verified observations are organized according to the general and specific requirements of the ISO/IEC 17025 standard. In this work, the nonconformities were classified into *major nonconformity*, *minor nonconformity*, *opportunity for improvement*, and in terms of the ECM, referred to as adverse events or errors.

The formal approach of the proposed errors reporting system in water sampling is based on the LP paradigm for problem-solving and set according to an extension to the LP language (Neves 1984). Undeniably, there are many approaches to Knowledge Representation and Reasoning (KRR) using the epitome of LP, namely, in the areas of Model Theory and Proof Theory. In this article, the Proof Theoretical approach for problem-solving was adopted and conveyed as an extension of the LP language. Indeed, an Extended Logic Program (ELP) is grounded on a specified ensemble of clauses in the form:

Program 1 The Archetype of an Extended Logic Program

$$\{$$

$$\neg\, p \leftarrow not\ p, not\ exception_p$$

$$p \leftarrow p_1, \cdots, p_n, not\ q_1, \cdots, not\ q_m$$

$$?\, (p_1, \cdots, p_n, not\ q_1, \cdots, not\ q_m)\ \ (n, m \geq 0)$$

$$exception_{p_1}, \quad \cdots \quad, exception_{p_j}\ \ (0 \leq j \leq k), \quad being\ k\ an\ integer\ number$$

$$\}$$

where p_i, q_j, and p stand for a set of predicates. A predicate is the part of a sentence that contains the verb and tells something about the subject. In program 1 ',' designates 'logical and' while '?' denotes 'falsity'. The classical negation sign '¬' denotes a strong declaration, whereas 'not' denotes negation by failure, i.e., a collapse in demonstrating a certain statement since it is not declared in an explicit way (Neves 1984). According to this way of thinking, a set of exceptions are presented in every program (Neves 1984), i.e., clauses in the form,

$$exception_{p_1}, \quad \cdots \quad, exception_{p_j}\ \ (0 \leq j \leq k), \quad being\ k\ an\ integer\ number$$

Table 1 | Categories of the adjusted ECM for water sampling and the corresponding codes

Category	Description	Code
Technical Dimension		
External	Technical faults outside the control and responsibility of the company.	TEX
Design	Faults caused by inadequate design of equipment, software, tags, or templates.	TD
Construction	Construction faults regardless of a right design.	TC
Materials	Material deficiencies not categorized as TD or TC.	TM
Reagents	Faults caused by alack or poor quality of chemicals.	TR
…	…	…
Organizational Dimension		
External	Faults at an organizational level outside the control of the company.	OEX
Transfer of Knowledge	Faults due to inadequate decisions that do not guarantee that the knowledge is shared with inexperienced employees.	OK
Protocols	Faults due to the quality/availability of the internal procedures (excessively intricate/simple, imprecise, or non-existent).	OP
Management Priorities	Internal choices in which safety is transferred to a lower position, giving priority to production.	OM
Culture	Faults due to the collective attitude towards risk and behaviors within the company.	OC
…	…	…
Human Behaviour Dimension		
External	Human faults that have its origins outside the control of the company.	HEX
Knowledge-Based Behaviour		
Knowledge-Based Errors	The disability of a member to put in practice current knowledge in a new situation.	HKK
…	…	…
Rule-Based Behaviour		
Qualifications	Inappropriate fit among the task to be performed and the employee's training, qualifications, or education.	HRQ
Coordination	Deficient task coordination (e.g., a fundamental task not fulfilled because the team was convinced that someone performed it).	HRC
Verification	Faults during the performing an accurate and full assessment of the set of circumstances before starting the execution of a specific task, covering the conditions and materials and/or equipment to be used.	HRV
Intervention	Faults that arise from defective preparation of a task and/or bad implementation.	HRI
Monitoring	Faults that occur during the monitoring of an activity or process.	HRM
…	…	…
Skill-Based Behaviour		
Slips	Faults in performing a task related to the operator's lack of fine motor skills.	HSS
…	…	…
Other		
Technicians' Related Factor	Faults associated with physical and/or psychic conditions of the employee that affect the task execution and are outside the control of the company.	TRF
Consumers' Related Factor	Faults associated with consumers' conduct that can affect water quality and are outside the control of the company.	CRF
…	…	…
Unclassifiable	Faults that cannot be included in the previous categories.	X

that denotes data, information, or knowledge that cannot be ruled out. On the other hand, clauses of the type viz.

$$?(p_1, \cdots, p_n, not\ q_1, \cdots, not\ q_m)\ \ (n,\ m \geq 0)$$

are invariants that make it possible to specify the context under which the universe of discourse should be understood (Neves 1984).

In order to isolate a customized version of the ECM for a particular problem, it is of the utmost importance to identify all the causes that lead to such an undesirable event. To adapt the ECM to the problems related to the sampling of water for human consumption, new dimensions have been considered. These dimensions aim to adjust the original model to the sampling context and provide a general view of the events that can happen. The Adjusted version of ECM for Water Sampling (A-ECM-WS) followed the dimensions *Technical*, *Organizational*, *Human Behaviour*, and *Other* presented in Table 1.

Table 1 also shows the categories included in each dimension, as well as a brief description and the corresponding codes. To illustrate the type of adverse events that can be included in each code, Table 2 presents a few examples. Thus, based on Tables 1 and 2, the systematization of the causes that lead to an adverse event becomes straightforward and effective, being processed according to the flow chart presented in Figure 2. For instance, an adverse event assigned to the category *Organizational Dimension – Protocols* (OP) occurs due to the quality/availability of the internal procedures. The events assigned to this category arise from difficulties in execution or interpretation of unclear protocols that can lead to adverse events such as the failure of not flowing the stagnant water in the faucet or the failure to remove the faucet accessories.

The ECM makes use of root cause analysis to classify adverse events through codes established in advance (van der Schaaf 1995). Thus, beyond the classification framework presented above, CTs regarding each adverse event should be conceived. CTs can be considered as a hierarchical representation of an adverse event, allowing to identify the circumstances that lead to its occurrence (van der Schaaf & Habraken 2005). The original version of the ECM only comprises cases for which the causes of adverse events are well known. This fact should be a limitation, since, in many situations, information about errors and adverse events can be incomplete, unknown, or even contradictory. To include the uncertainty that is always present in real situations, in the present work, the *unknown* and *forbidden* operators were considered in the conception of the extended causal trees (ECT).

APPLICATION

Table 3 lists the adverse events and the number of occurrences recorded in the internal audits carried out in the Water Laboratory of Santiago do Cacém Municipality (located in Alentejo Litoral, southern Portugal) during the 1st semester of 2019. A perusal of Table 3 reveals that 128 adverse events occurred. Considering the general requirements of the ISO/IEC 17025 standard, 28.9% of the occurrences were related *Resource Requirements*, 53.9% were associated with *Processes Requirements*,

Table 2 | An excerpt from the code set of the adjusted ECM for water sampling and examples of possible adverse events

| Code	Extension to water sampling	Examples
TR	Faults caused by a lack or poor quality of chemicals.	Use reagent out of date
OP	Very complicated, inadequate, or missing protocols. Protocols not disclosed to the team.	Nonconformity of protocols with various negative effects (e.g., non-flow of stagnant water in the duct section, do not remove accessories outside the tap).
HEX	Human faults outside the control of the company.	Excessive time between sampling and delivery to the laboratory.
HKK	Do not put on preservatives or choose the wrong preservative.	Do not add HNO_3 in the samples for determining iron and manganese.
HRI	Failure in task planning or during its execution.	Did not disinfect hands before sampling, did not use gloves, did not use the blowtorch correctly, and did not choose a tap under normal conditions of conservation and hygiene.
HRM	Faults that occurred during the monitoring of a task.	Failure to calibrate the data logger used inside the thermal bags.
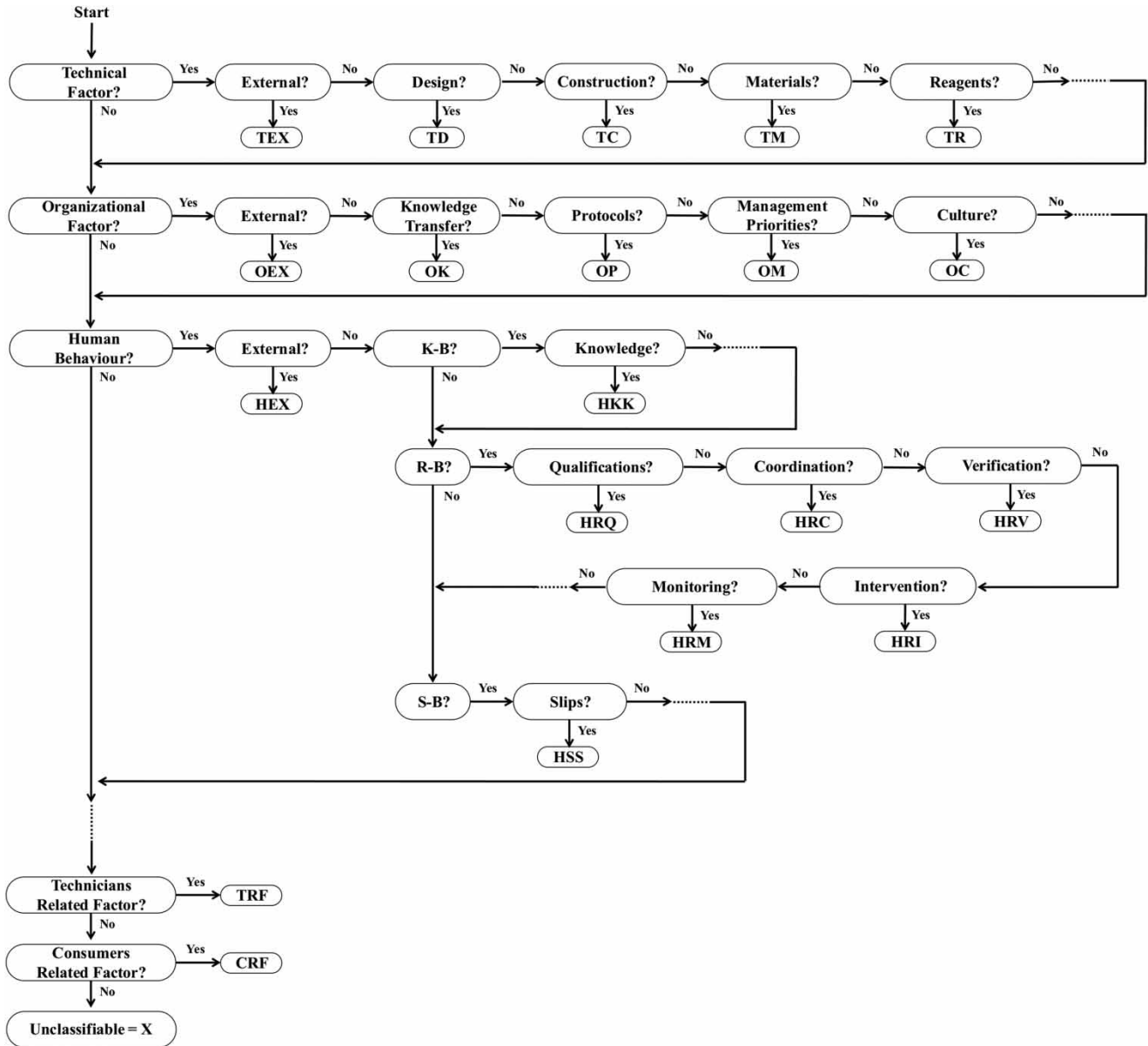
Figure 2 | The adverse event classification process.

whereas 17.2% were linked to *Management System Requirements*. In terms of nonconformities, 81.2% were classified as *minor nonconformity* (*N*) and 18.8% corresponded to *major nonconformity* (*M*).

Considering the data presented in Table 3, it can be argued that the sample-related adverse events occur much more frequently than the others, which require special attention. From the nonconformities recorded during the 1st semester of 2019, it is feasible to specify the actions and causes that contribute to the occurrence of the adverse event 'Failure in Sampling'. The ECT presented in Figure 3 was established with the collaboration of sampling technicians. The ECT was the subject of an in-depth study before its final structure took shape. The ECT contains feasible causes, being a symbolic depiction of a problem. Each branch of the ECT describes the causes of a particular adverse event.

To illustrate the operating mode of the A-ECM-WS, the undesired event 'Error in Sampling' and the respective causal tree (Figure 3) were considered. Taking into account the fact that an adverse event can emerge due to several causes that should be considered simultaneously or separately, *AND/OR-nodes* were incorporated in the CT. Furthermore, the operators 'unknown' and 'forbidden' were also added to express events for which the causes are unknown, forbidden, or not permitted. Therefore, taking into account the information available in Figure 3, the following different scenarios can be examined:

Table 3 | Number of adverse events that occurred during the 1st semester of 2019

General requirements	Specific requirements	Description	Classification	Number of occurrences	Total
Resource requirements	Facilities and environmental conditions	Laboratory facilities are not adequate (e.g., no separation between incompatible testing areas).	M	2	7
		The laboratory did not formalize the facilities and environmental conditions.	N	5	
	Metrological traceability	The laboratory did not formalize the methodology for estimating uncertainties.	N	7	22
		The calculation of the uncertainty estimate for the temperature test was not shown.	N	9	
		The calculation of the uncertainty estimate for free and total residual chlorine was not shown.	N	6	
	Equipment	The laboratory did not show records of annual visual inspections of the volumetric glass material.	N	8	8
Process requirements	Nonconforming work	The methodology for nonconforming work is not formalized.	M	10	10
	Selection, verification, and validation of methods	The laboratory has not formalized the requirement methods/procedures for testing and validating the methods.	N	5	9
		The laboratory did not formalize the methodology for the implementation of the test methods.	N	4	
	Sampling	Sampling technician failures were found (e.g., failure in disinfection).	N	12	50
		It was found that in the procedure there were deficiencies in the location of the sampling point.	N	6	
		It was found that inappropriate bottles were being used to carry out the sampling.	N	12	
		Failures were found in the preservation of samples.	N	10	
		Transport failures were noted (e.g., agitation during transport, time between collection, transport, and testing)	N	10	
Management system requirements	Documentation	Documentation control is neither adequate nor formalized.	M	12	22
		The external documents to be controlled are not defined.	N	10	
Adverse events registered					128

M, major nonconformity; N, minor nonconformity.

- Scenario A – The 'Failure in Sampling' was due to microbiological contamination caused by the non-removal of the faucet accessories.
- Scenario B – The 'Failure in Sampling' was due to chemical contamination caused by a failure in sample preservation.
- Scenario C – A 'Failure in Sampling' has occurred. The causes behind the event were not recorded, and
- Scenario D – A 'Failure in Sampling' has occurred. The causes of the adverse event were not recorded due to the internal rules of the company.

In scenario A, the cause of the adverse event was recorded, corresponding to a known value. Scenario B corresponds to a situation in which the cause that led to the adverse event is unknown. However, it is possible to lay down a finite set of causes that can lead to its occurrence, configuring a situation of an unknown value in a finite set of values (two in this example, wrong preservative or reagent out of data). In scenario C, the occurrence of the adverse event was recorded, but nothing

is known about their causes, corresponding to an unknown value. Finally, scenario D involves a forbidden or a not-allowed value, since it is not permitted to know the causes of the adverse event.

The four scenarios described above can be depicted in Program 2, following the archetype presented in Program 1, in terms of the extent of predicate **adverse_event_cause** (*aec*), namely,

Program 2 Structure of the logic program in relation to scenarios A, B, C, and D.

{

{

/* The following sentence denotes that the extent of the predicate $aec_{scenario\ A}$ is based on the explicitly specified clauses and those that cannot be dropped, i.e., the exceptions */

$$\neg\ aec_{scenario\ A}\ (X) \leftarrow not\ aec_{scenario\ A}\ (X),$$
$$not\ exception_{aec_{scenario\ A}}\ (X).$$

/* The following sentence stands for an axiom of $aec_{scenario\ A}$ */

$aec_{scenario\ A}\ (failure\ in\ desinfection\ of\ faucet).$

}

{

/* The following sentence denotes that the extent of the predicate $aec_{scenario\ B}$ is based on the explicitly specified clauses and those that cannot be dropped, i.e., the exceptions */

$$\neg\ aec_{scenario\ B}\ (X) \leftarrow not\ aec_{scenario\ B}\ (X),$$
$$not\ exception_{aec_{scenario\ B}}\ (X).$$

/* The following sentence stands for adbucible axioms of $aec_{scenario\ B}$ */

$exception_{aec_{scenario\ B}}\ (wrong\ preservative).$

$exception_{aec_{scenario\ B}}\ (reagent\ out\ of\ data).$

}

{

/* The following sentence denotes that the extent of the predicate $aec_{scenario\ C}$ is based on the explicitly specified clauses and those that cannot be dropped, i.e., the exceptions */

$$\neg\ aec_{scenario\ C}\ (X) \leftarrow not\ aec_{scenario\ C}\ (X),$$
$$not\ exception_{aec_{scenario\ C}}\ (X).$$

/* The following sentence stands for an adbucible axiom of $aec_{scenario\ C}$ */

$exception_{aec_{scenario\ C}}\ (unknow).$

}

{

/* The following sentence denotes that the extent of the predicate $aec_{scenario\ D}$ is based on the explicitly specified clauses and those that cannot be dropped, i.e., the exceptions */

$$\neg\, aec_{scenario\ D}\,(X) \leftarrow not\ aec_{scenario\ D}\,(X),$$
$$not\ exception_{aec_{scenario\ D}}\,(X).$$

/* The following sentence stands for an invariant with respect to predicate $aec_{scenario\ D}$ */

$$?\ aec_{scenario\ D}\,(X), not\ null_{scenario\ D}(X).$$

/* The following sentence denotes that the extent of the predicate null is based on the explicitly specified clauses and those that cannot be dropped, i.e., the exceptions */

$$\neg\, null_{scenario\ D}\,(X) \leftarrow not\ null_{scenario\ D}\,(X),$$
$$not\ exception_{null_{scenario\ D}}\,(X).$$

/* The following sentence stands for an axiom of null's predicate */

$$null_{scenario\ D}\,(forbidden).$$
$$\}$$
$$\}$$

In decision-making processes, it is necessary to consider only the most promising answers. Thus, each answer must be tested in terms of its ability to adapt to dynamic environments (i.e., environments that are constantly updated) to choose the most appropriate action from a wide range of alternatives. The optimal response, in terms of the LP paradigm, is a logic program or theory that models the universe of discourse, maximizing the Quality of Information (QOI; Vilhena *et al.* 2016; Coelho *et al.* 2017). The *QoI* with respect to a generic predicate p can be analyzed in four situations and can be measured in the interval [0, 1]. With regard to cases in which information is known, the *QoI* for the extension of predicate

Figure 3 | The ECT regarding the adverse event 'Failure in Sampling'.

p is 1. For cases where information is unknown, the *QoI* is given by

$$QoI_p = \lim_{N \to \infty} \frac{1}{N} = 0, \quad (N \gg 0) \tag{1}$$

where N stands for the number of terms or clauses of the extension of predicate p. For cases in which information is unknown but can be selected from a set of possible values, although only one can be considered, the *QoI* is given by

$$QoI_p = \frac{1}{Card} \tag{2}$$

with *Card* being the cardinality of the set of exceptions of p. If more than one value can be considered, the *QoI* is given by

$$QoI_p = \frac{1}{(C_1^{Card} + C_2^{Card} + \cdots + C_{Card}^{Card})} \tag{3}$$

where C_{Card}^{Card} is a card-combination subset, with *Card* elements (Vilhena *et al.* 2016; Coelho *et al.* 2017).

The information included in the ECT was depicted in terms of the LP paradigm through the extensions of a predicate set (Program 2), which make possible to evaluate the *QoI* regarding each of the scenarios described above. In the case of scenario A, there is no doubt about the cause of the adverse event. The sampling failure was caused by the non-removal of the faucet accessories, with *QoI* being equal to 1. In scenario B, the cardinality of the set of exceptions of the predicate is 2, i.e., the usage of a wrong preservative or an out-of-data reagent, with *QoI* being equal to 1/2. Finally, in scenarios C and D, it is known that the adverse event took place, but its causes are unavailable. Any value presented in the ECT can be assumed as a feasible solution and the *QoI* is 0.

Using the formal approach based on the LP paradigm mentioned above, an AERLS-WS was developed. In addition to recording adverse events, the AERLS-WS is an information management system that provides a knowledge base where it is possible to use data mining tools. Thus, the AERLS-WS includes three main modules as shown in Figure 4. The Reporting Forms Module (RFM) allocates the web interface that allows the registration of adverse events. The recording can be carried out by any laboratory employee, regardless of the department, by way of predefined forms adapted to each user profile (e.g., sampling technicians, chemical/microbiological technicians, technical managers, and management team). Bearing in mind what just been mentioned, the focus in the development of this interface was its usability.

The Manager Reports Module (MRM) is equally a web interface. It intends to make possible analysis of the adverse events previously recorded in the RFM, having the A-ECM-WS subjacent. The A-ECM-WS allows that the system provides suggestions for improvements to minimize or eliminate organization failures. Thus, the system can issue automatically individual reports for the adverse events recorded, as well as details in terms of the branches of the ECT. The improvement suggestions contained in the reports are due to the methodology followed that combines an adjusted version of the ECM for water sampling and LP for knowledge representation and supporting the reasoning mechanisms. The system's reports are essential to ensure quality to laboratory results, since the time and resources required to study the causes of the events, to make inferences, and to establish patterns and trends are substantially reduced. In this way, this task (which is usually time-consuming) can be performed speedily, contributing to avoid/mitigate failures. In addition, once the data are stored in the same database, they are always available for further studies. The MRM can provide updated information about the impact, occurrence, and type of events recorded. The user can choose how the information is displayed (e.g., bar charts, pie charts, tables, and tree diagrams). Based on the system's reports, the user can have information about the contribution of each possible action/ cause to the adverse event. For example, for the adverse event corresponding to scenario A (i.e., failure in sampling due to microbiological contamination caused by the non-removal of the faucet accessories), the report generated by the system informs how many times it occurred, the places and dates of the occurrences, and the error consequences (e.g., invalid analysis) and recommends remedial actions to minimize the problem (e.g., the report suggests that sampling technicians should pay careful attention to sampling rules for microbiological analysis). Moreover, based on the reports provided by the system, the management team can assess the need to provide training to sampling technicians. It should be noted that the actions/causes were identified by the system automatically, considering the A-ECM-WS and the ECTs, contributing to avoid or overcome the identified issues and errors.
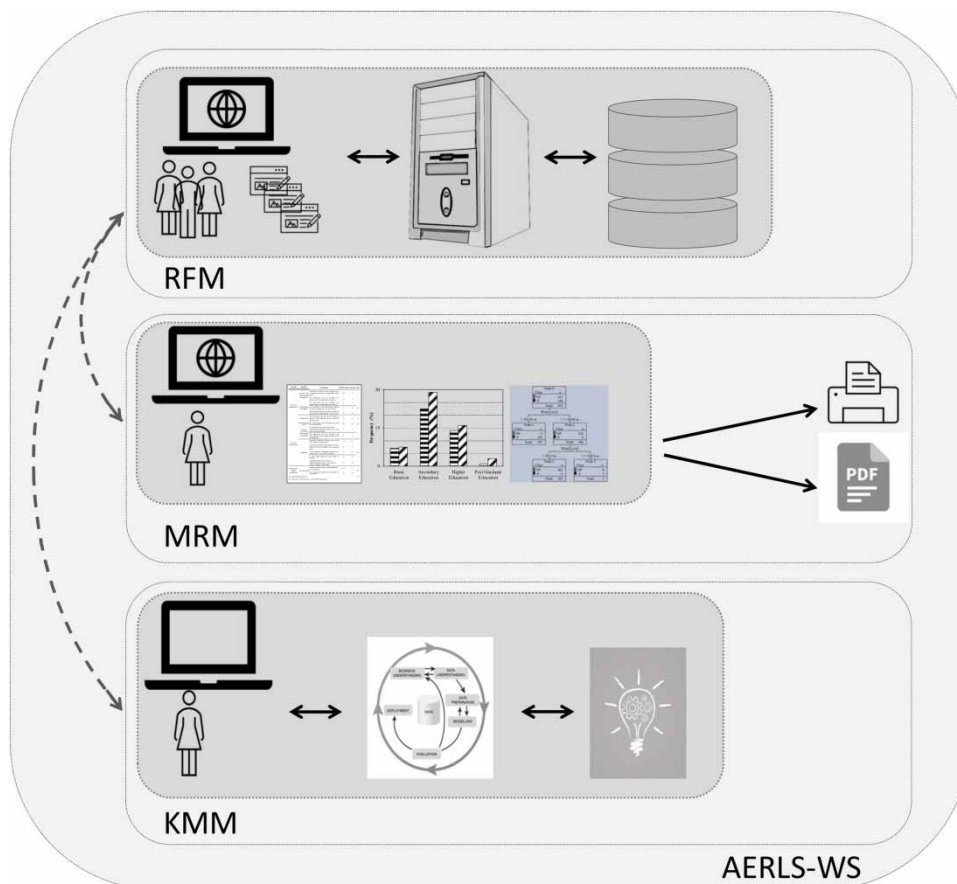
Figure 4 | A global view of the Adverse Event Reporting and Learning System related to Water Sampling (AERLS-WS).

Lastly, the Knowledge Manager Module (KMM) is a Java application. This module makes use of the records stored in the database to generate a knowledge base from which other relevant studies (e.g., through the use of data mining tools) aiming to optimize the sampling procedures can be carried out.

CONCLUSIONS

To achieve accurate results in water analysis, it is essential that sampling is carried out correctly. To help fulfill this objective, this work presents an adverse event reporting and learning system, developed specifically for the sampling phase, that aims to improve the quality of drinking water and, therefore, contributes to the promotion of public health. The key contribution of this work is related to the quantification of the QoI imbedded in CT and to address the issues pertaining to incomplete information using the LP paradigm. The proposed approach allows to identify in advance the causes that lead to adverse events, enabling the formulation of action plans to avoid or, at least, minimize them. The construction of a CT for a particular event is often time-consuming and difficult, because initially it requires the development of branches for each possible cause. However, after this stage is passed, the process of classification and analysis of adverse events becomes a speedier and easy exercise. A computation of the *QoI* enables the ranking of possible causes and to identify the ones that should be prioritized. Consequently, in the construction of a CT, it is mandatory to look for all possible causes, not only the most probable or the usual ones.

The adverse event reporting and learning system presented allows knowledge extraction, i.e., the identification of the main failure causes, possible trends, areas requiring improvement plans, or changes in procedures. The automatic reports generated by the system are reliable. In fact, the quantification on-the-fly of the *QoI* of the logic terms used in the process of judgement (given in terms of a theorem to be proved) minimizes user participation, which is only necessary to approve the

recommendations. One of the main advantages of the proposed system has to do with its modularity. This feature allows one to update the system at any time (e.g., add new categories, subcategories, and/or causal trees) without changing its structure or way of functioning.

ACKNOWLEDGEMENTS

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

Allan, R., Jeffrey, P., Clarke, M. & Pollard, S. 2013 The impact of regulation, ownership and business culture on managing corporate risk within the water industry. *Water Policy* **15**, 458–478. https://doi.org/10.2166/wp.2013.175.

Bagchi, B. 2013 *Water in Biological and Chemical Processes*. Cambridge University Press, Cambridge.

Brock, T. 1999 *Robert Koch: A Life in Medicine and Bacteriology*. American Society for Microbiology, Washington.

Byleveld, P. M., Deere, D. & Davison, A. 2008 Water safety plans: planning for adverse events and communicating with consumers. *Journal of Water and Health* **6** (S1), 1–9. https://doi.org/10.2166/wh.2008.035.

Coelho, C., Vicente, H., Martins, M. R., Lima, N., Neves, M. & Neves, J. 2017 An assessment of environmental and toxicological risk to pesticide exposure based on a case-based approach to computing. *IOP Conference Series: Earth and Environmental Science* **52**, 012091. https://doi.org/10.1088/1742-6596/52/1/012091.

Eaton, A., Clesceri, L., Rice, E. & Greenberg, A. 2017 *Standard Methods for the Examination of Water and Wastewater*, 23rd edn. American Public Health Association/Water Works Association/Water Environment Federation, New York.

Fernandes, A., Carvalho, F., Ribeiro, J., Vicente, D., Faria, J., Figueiredo, M., Capita, A., Neves, J. & Vicente, H. 2019 Fully informed classification systems simpler, maybe better. *Communications in Computer and Information Science* **1038**, 3–16. https://doi.org/10. 1007/978-3-030-37858-5_1.

Fernandes, A., Lima, R., Figueiredo, M., Ribeiro, J., Neves, J. & Vicente, H. 2020 Assessing employee satisfaction in the context of COVID-19 pandemic. *ParadigmPlus* **1** (3), 23–43.

Geison, G. 1995 *The Private Science of Louis Pasteur*. Princeton University Press, Princeton.

Henneman, E. A., Blank, F. S. J., Gattasso, S., Williamson, K. & Henneman, P. L. 2006 Testing a classification model for emergency department errors. *Journal of Advanced Nursing* **55**, 90–99. https://doi.org/10.1111/j.1365-2648.2006.03878.x.

Hommerson, A., Lucas, P. & van Bommel, P. 2008 Checking the quality of clinical guidelines using automated reasoning tools. *Theory and Practice of Logic Programming* **8**, 611–641. https://doi.org/10.1017/S1471068408003451.

ISO/IEC 17025:2017 2017 *General Requirements for the Competence to Testing and Calibration Laboratories*. International Organization for Standardization, Geneva.

Kakas, A., Kowalski, R. & Toni, F. 1998 The role of abduction in logic programming. In: *Handbook of Logic in Artificial Intelligence and Logic Programming*, Vol. 5 (Gabbay, D., Hogger, C. & Robinson, I., eds). Oxford University Press, Oxford, pp. 235–324.

Keet, C. M. 2013 Open World Assumption. In: *Encyclopedia of Systems Biology* (Dubitzky, W., Wolkenhauer, O., Yokota, H. & Cho, K.-H., eds). Springer-Verlag, New York, p. 1567.

Li, Y. & Migliaccio, K. 2011 *Water Quality Concepts, Sampling and Analyses*. CRC Press, Boca Raton.

Morales-Torres, A., Escuder-Bueno, I., Andrés-Doménech, I. & Perales-Momparler, S. 2016 Decision support tool for energy-efficient, sustainable and integrated urban stormwater management. *Environmental Modelling & Software* **84**, 518–528. https://doi.org/10.1016/j. envsoft.2016.07.019.

Neves, J. 1984 A logic interpreter to handle time and negation in logic data bases. In: *Proceedings of the 1984 Annual Conference of the ACM on the Fifth-Generation Challenge* (Muller, R. L. & Pottmyer, J. J., eds). Association for Computing Machinery, New York, pp. 50–54.

Neves, J., Maia, N., Marreiros, G., Neves, M., Fernandes, A., Ribeiro, J., Araújo, I., Araújo, N., Ávidos, L., Ferraz, F., Capita, A., Lori, N., Alves, V. & Vicente, H. 2021 Employees balance and stability as key points in organizational performance. *Logic Journal of the IGPL*. https:// doi.org/10.1093/jigpal/jzab010.

Nollet, L. & Gelder, L. 2014 *Handbook of Water Analysis*, 3rd edn. CRC Press, Boca Raton.

Parsons, S. 1996 Current approaches to handling imperfect information in data and knowledge bases. *IEEE Transactions on Knowledge and Data Engineering* **8**, 353–372. https://doi.org/10.1109/69.506705.

Pearson, L., Coggan, A., Proctor, W. & Smith, T. 2010 A sustainable decision support framework for urban water management. *Water Resources Management* **24**, 363–376. https://doi.org/10.1007/s11269-009-9450-1.

Raab, S. S., Vrbin, C. M., Grzybicki, D. M., Sudilovsky, D., Balassanian, R., Zarbo, R. J. & Meier, F. A. 2006 Errors in thyroid gland fine-needle aspiration. *American Journal of Clinical Pathology* **125**, 873–882. https://doi.org/10.1309/7rqe-37k6-439t-4pb4.

Rasmussen, J. 1976 Outlines of a hybrid model of the process plant operator. In: *Monitoring Behavior and Supervisory Control*, Vol. 1 (Sheridan, T. B. & Johannsen, G., eds), NATO Conference Series. Springer, Boston, pp. 371–383. https://doi.org/10.1007/978-1-4684-2523-9_31.

Reason, J. 2000 Human error: models and management. *BMJ* **320** (7237), 768–770. https://dx.doi.org/10.1136/bmj.320.7237.768.

Rodrigues, S., Brandão, P., Nelas, L., Neves, J. & Alves, V. 2011 A logic programming approach to medical errors in imaging. *International Journal of Medical Informatics* **80**, 669–679. https://doi.org/10.1016/j.ijmedinf.2011.06.005.

Simmons, D. & Graves, K. 2008 Tubing misconnections – a systems failure with human factors: lessons for nursing practice. *Urologic Nursing* **28**, 460–464.

Snow, J. 1855 *On the Mode of Communication of Cholera*. John Churchill, London.

Stefanovic, N., Radojevic, I., Ostojic, A., Comic, L. & Topuzovic, M. 2015 Composite web information system for management of water resources. *Water Resources Management* **29**, 2285–2301. https://doi.org/10.1007/s11269-015-0941-y.

van der Schaaf, T. W. 1995 Near miss reporting in the chemical process industry: an overview. *Microelectronics Reliability* **35** (9–10), 1233–1243. https://doi.org/10.1016/0026-2714(95)99374-R.

van der Schaaf, T. W. & Habraken, M. M. P. 2005 *PRISMA-Medical: A Brief Description*. Patient Safety Systems, Faculty of Technology Management, Eindhoven University of Technology, Eindhoven.

Vicente, H., Borralho, F., Couto, C., Gomes, G., Alves, V. & Neves, J. 2015 An adverse event reporting and learning system for water sector based on an extension of the Eindhoven Classification Model. *Water Resources Management* **29**, 4927–4943. https://doi.org/10.1007/s11269-015-1075-y.

Vilhena, J., Vicente, H., Martins, M. R., Grañeda, J., Caldeira, F., Gusmão, R., Neves, J. & Neves, J. 2016 A case-based reasoning view of thrombophilia risk. *Journal of Biomedical Informatics* **62**, 265–275. https://doi.org/10.1016/j.jbi.2016.07.013.