Formalising the Language of Risk

The use of natural language in engineering and specifically, engineering risk management, is often problematic due to assumed meanings and usage contexts of domain terms, with the result that misunderstandings can arise. Dave Banham provides an introduction to a formalised structure of words – an ontology – by which, at least, the risks arising from system safety and security concerns can be described unambiguously using a common language.

The Problem

The Data Safety Initiative Working Group (DSIWG) has been working on guidance for the management of data safety risks for several years and has now produced mature guidance material. The group has however, found that the language of safety and risk is contextually dependent with terms having multiple meanings, different terms being used with the same implied meaning, and, as is often the case in English, terms being used without qualification.

Add to this situation the risk terminology used by cyber-security professionals, because security informed data safety is a useful adjunct, and the result is a long glossary of terms with no self-consistency. It is therefore hard to produce guidance for data safety that is clear and accessible. Moreover, often this variability of meaning, introduces subtle misunderstandings that take conscious effort to detect and resolve.

To help resolve these issues the DSIWG established a sub-working group to establish a formalised structure of words – an ontology - to describe risk. The intention was to use the ontology to be more precise in the language used in the guidance, but it has been realised this ontology could have much wider applicability across the entire safety and security domains.

What do we mean by Risk?

In common usage, "risk" means an activity or situation that has a chance of a significantly unpleasant outcome in the worldview of the observer making the statement. Risk is generally used as the adjective "risky" to qualify the sense of uncertainty being expressed about the named activity: *parachute diving is risky; driving fast is risky; betting on slot machines is risky*, etc. Risk can be used as a noun when conceptualising it: *I accept the risks involved in free climbing; the risk of injury in rugby is high*. The two forms can be combined to yield sentences such as: *there is a risk of injury from risky driving*. However, note how easily the implied meaning of risk shifts from one of uncertainty in outcome (when used as adjective), to that of likelihood (when used as noun).

Moreover, the outcomes and likelihoods that are often inexplicitly stated as an assumed shared understanding, are significantly undesirable in the worldview of the person making the statement, but may be considered otherwise by somebody else. Free climbing is one person's *horror story*, but another's *pleasant sport*. The common language use of the word "risk" is completely inadequate for engineering where assumptions need to be eliminated in preference for precise terms, calculations, and a shared (and agreed) worldview.

Articulating Risk

Engineering makes use of standardised terms to help articulate risk. The international standard for risk management is ISO 31000 [1], with the compendium ISO vocabulary of risk terms, ISO Guide 73 [2].

ISO Guide 73 defines risk as the *effect of uncertainty on objectives of stakeholders*. Objectives are things that stakeholders seek or want to avoid. We don't want harm to arise from the use of our goods and services; conversely, we want to make money from selling our goods and services. Uncertainty exists in the fulfilment of these objectives due to phenomena such as natural processes, unforeseen circumstances, competition, etc. When things are certain (perhaps because they have already happened), there is no risk.

This leads to the idea of positive risk (*seeking a benefit*) and negative risk *(avoiding a harm*). The common use of "risk" is in the sense of negative risk; harmful (often physical) situations that need to be avoided. However, risk arises from the worldview, frame, or context that the stakeholder has since they own the objective. Consider theft. The owner of a valuable asset wants to protect that asset from, amongst other things, theft. Theft results in a harm that creates a loss, to the owner, of the stolen asset and is thus a negative risk concept to the stolen asset's owner. Whereas to the criminal, theft is the means by which value is gained (a benefit) and is thus a positive risk concept to them, notwithstanding the negative risk of being caught.

The language of safety is formalised around that of risk. Let us start by defining "harm". Harm is the *consequence of a failure* to meet stakeholder objectives when the consequential situation is undesirable to them. The converse is a "Benefit", when the *consequence is desirable* to them. A subset of the total set of possible harms is the set of safety-related harms. A safety-related harm is generally defined as a physical harm that impacts the health or life of a person or persons, or impacts the wellbeing of the natural environment. Although a stakeholder may include other impacts such as the loss of an asset, loss of reputation, etc. in their definition.

How can harms arise? Since harms are generally not certain, specific situations need to occur to allow them to arise as a *consequence*. These causal situations are referred to as *incidents*. An incident is a *dangerous event* (i.e. a moment in time) and is therefore a *danger source*. (That is, danger may lead to harm.)

A near miss is an incident that did not lead to harm, but had the potential to do so. An accident arises from an unintentional incident that leads to harm; that is, an accident arises from unintentional sources of danger.

Incidents can be intentionally created and sometimes maliciously; for example, by arsonists, thieves, or by misguided misuses of a system (i.e. by incompetent users). As such, the term "incident" is more useful than purely safety terms such as "accident", as it allows the safety analysis to consider a wider set of concerns that have traditionally been, for example, the reserve of security specialists.

One purpose of a system safety analysis is to theorise about what potential harms a system may cause and to identify the potential danger sources that may lead to them. An identified danger source is called a *hazard*; a hazard is a known danger source that may lead to an incident that causes harm.

A risk score is a metric arising from a function of a potential incident's likelihood and the desirability of the potential outcome. Hence, numerically:

risk = likelihood × desirability

where *likelihood* is a probability of occurrence, *desirability* is a positive score when the objective is sought and a negative score when it is to be avoided, and where × is a binary operator (i.e. a function) taking two parameters. Hence, a negative risk score indicates the risk of a harm, which conforms to the ISO 31000 framework. In the context of harms, desirability is often stated as a severity score and the equation of negative risk can be stated as:

risk = -(likelihood × severity)

To understand how harm may arise we either start with a harm and ask the deductive question of *how can it arise*, or we start with some other aspect of the system such as a system input, or a subsystem and ask the inductive question of *what would happen if*. Where a weakness or susceptibility to failure is found in a constituent part of a system (which also includes people when they form an active part of the system) then a *vulnerability* is said to exist.

A danger source is something that can exploit a vulnerability to create an incident. An incident that is a system failure can result in harm, although typically what happens is that the incident is a failure that is more localised to a constituent part of the system; that is, a part no longer completely fulfils its objectives. A localised failure manifests as a fault (failure condition) that can propagate through a system exploiting other vulnerabilities (that is, triggering other failures) until potentially the system fails with some harmful outcome.

A Model of Risk Terminology

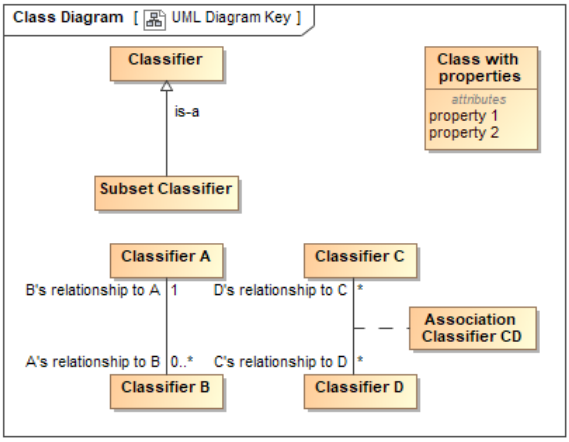
We can describe this terminology formally in an ontology and use UML class diagrams to represent aspects of that ontology through a series of diagrams. The ontology captures the ISO 31000 concept of desired and undesired stakeholder objectives through the consequences of benefit and harm. However, from a safety and security point of view, our main interest is in the risks associated with harms and, as a result, the ontology is significantly more refined in this area. Nevertheless, it is important to understand the opportunity and benefit cases that malicious threat actors may have towards a system.

The figure on the next page shows the graphical notation subset of UML 2 [3] class diagrams that are used to model the risk ontology. The rectangular shapes within the diagram frame represent classifiers, which are used to describe the language terms in the ontology.

The *is-a* relationship denoted by the hollow closed arrow headed edge (▷) describes a taxonomical relationship between classifiers where the specialised classifier is a subset concept of the more general classifier at the arrowhead end of the edge. The ontology optionally clarifies the specialisation with an annotation text shown with double angle quotes next to the generalisation edge as follows: «Classifies» and «Subsets»

A «Classifies» relationship denotes a set of specialisations that can be used in an additive fashion; that is, they are overlapping and additive concepts. Whereas a «Subsets» relationship denotes a set of specialisations that are distinct from each other; that is, they are non-overlapping concepts.

For example, a vehicle can be classified by its means of its source of power, its means of motion, and its colour (to name just a few). Each of these classifiers can be subset, so for example for power, we could have diesel engine, electric engine, gas turbine, etc., and for the subsets of means of motion we could have, wheels, wings, hull, etc., and for colour some set of colours.



From this set of classifiers and their subsets we can describe a vehicle as red, with diesel engine and wheels, or as red with gas turbine and wings. The classifiers are additive and individually describe an aspect of the thing (a vehicle in this example) they classify. They also provide discrimination by class, so in this example all the red vehicles can be identified, irrespective of their other classifiers.

A class can relate to other classes in non-taxonomical ways and these are denoted by edges with either no arrowheads, where the relationship is bidirectional (as shown in the figure for classifiers A and B), or with a single open arrowhead end (→) to denote a unidirectional relationship. The meaning of the relationship is denoted by the verb phrases at each end of the edge for bidirectional relationship, or just at the arrow headed end for unidirectional relationship.

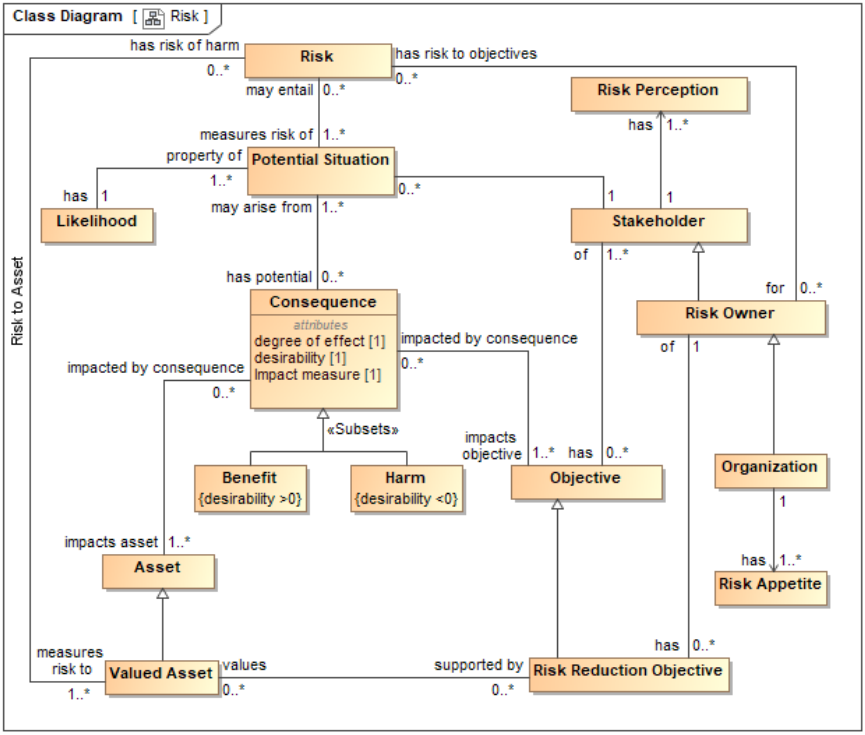
A relation end multiplicity quantifies the permissible number of relationships that may exist when the related terms are being used. The table below lists the typical multiplicities that have been used and their corresponding meaning. The combination of the multiplicity and the associated end verb phrase combine to provide a quantified relation from one classifier to the related classified.

Multiplicity designation	Meaning
1	One
0..1	May have (i.e. none or one)
1..*	Some (i.e. one or more)
*	May have some (i.e. none, one, or more)

An association can be further qualified by an association class (as shown in the figure for classifiers C and D with association class CD). The purpose of an association class is to provide a class based definition of the association; that is, an association class is a class with edges. One benefit this provides in ontology modelling is that it allows relationships to be defined terms by virtue of the association class name.

A Model for Risk

To show an example of the ontology, consider the following figure that shows the relationships associated with the entity "Risk".



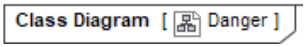
From the "Risk" Class Diagram on the previous page, we can derive the following narrative:

Risk is a measure of the uncertainty in attaining **stakeholder** objectives. **Objectives** are things that **stakeholders** want or do not want to happen, which are not certainties.

Risks are contextualised against the **asset** or assets that are impacted by the **objectives**, given the *likelihood* of the **potential situations** that may arise from them and the *desirability* of the **situation** that may arise as **consequence**. Such **consequences** can be *desired* or not *desired* and we call this a **benefit** and a **harm** respectively.

In terms of managing risk, there needs to be an identified **risk owner** that has **risk reduction objectives** that relate to the subset of **assets** that are considered to be of value (i.e. **valued assets**). This corresponds with the pragmatic view that the formulation of a risk treatment strategy needs to be targeted to be cost effective.

A further example of the ontology is given in "Danger" class diagram in the figure on the opposite page. In this figure we start to see how the ontology provides a language that is common between safety and security domains.



As with the previous figure, we can derive a narrative from the diagram opposite as follows:

Danger is described by the *possibility* that an **undesirable situation** may cause **harm**, as denoted by the relationship between these two terms. More specifically, we can state that **harm** arises from **incidents** that cause it; an **incident** is the cause and **harm** is the **consequence**.

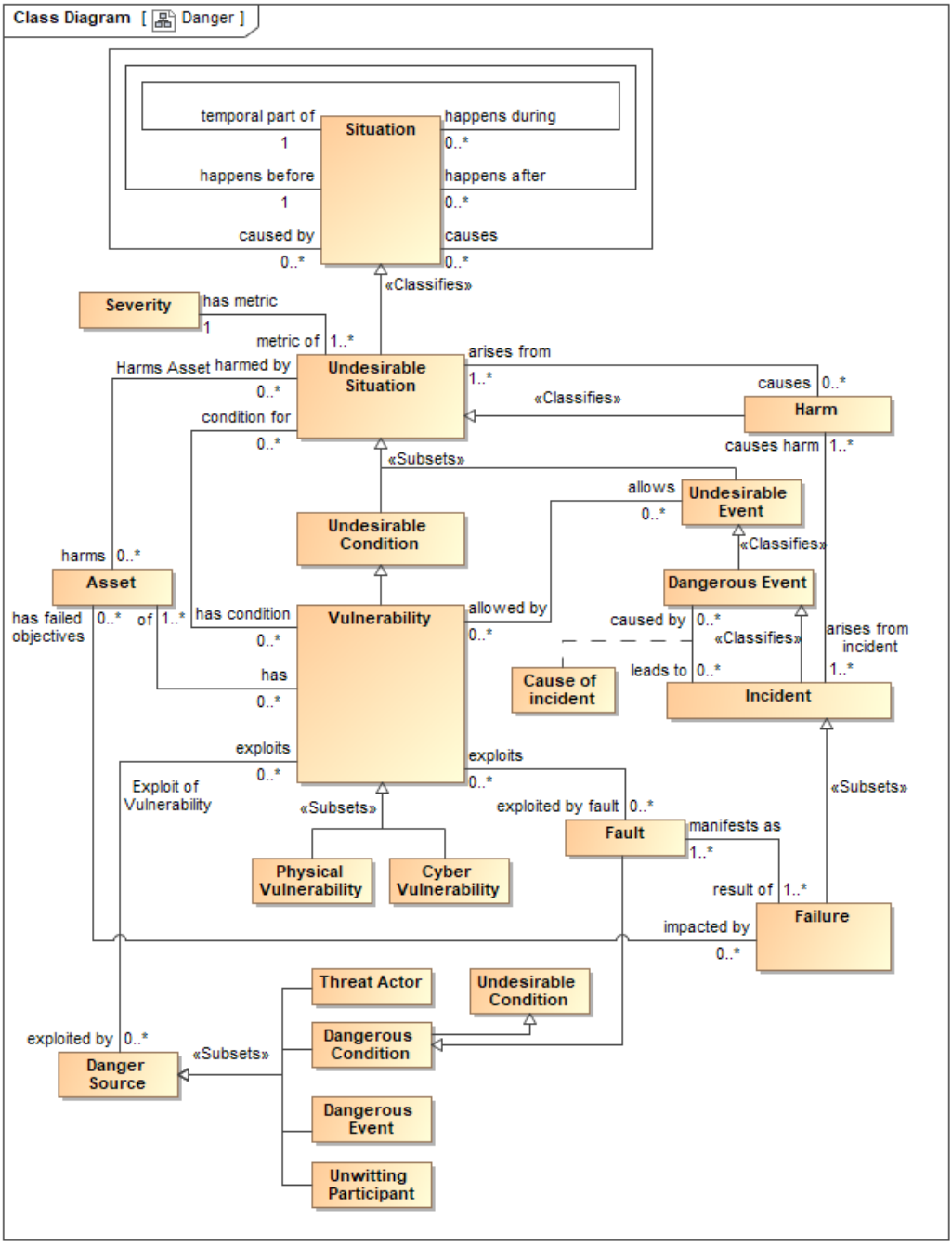
An **incident** is both a **danger source** (a **dangerous event**) and an **undesirable situation** (an **undesirable event**). Hence, the *possible* relationship between **undesirable situation** and **harm** becomes a substantiated one between **incident** and **harm**.

The degree of *danger* posed by an **undesirable situation** is captured by its **severity** property.

Assets can have **vulnerabilities**, where a **vulnerability** defines the conditions that *allow* an **undesirable event** to occur. A **danger source** is the generic term for something – natural, systematic, or intentional – that can *exploit* a **vulnerability** to create **an undesirable event**.

Since **undesirable events** can be classified as **dangerous events**, which is a **danger source**, a chain of events can be created whereby a series of **vulnerabilities** are exploited until an **incident** occurs and **harm** arises.

The term "exploit" is used here with both its "use" and "abuse" meanings. Physical things have **physical vulnerabilities** that are subject to the laws of physics and particularly the law of entropy; physical things break. They break through the *wear and tear* of natural use, and they break by being abused and misused. Complex systems have both physical vulnerabilities and vulnerabilities arising out of design limitations and design flaws (i.e. systematic defects). In computer-based systems, these design vulnerabilities are called **cyber vulnerabilities.**


Class Diagram [Danger]
temporal part of
1
Situation
happens during
0..*
happens before
1
happens after
0..*
caused by
0..*
causes
0..*
«Classifies»
Severity
has metric
1
metric of
1..*
Harms Asset
harmed by
0..*
Undesirable Situation
arises from
1..*
«Classifies»
causes
0..*
Harm
condition for
0..*
causes harm
1..*
«Subsets»
allows
0..*
Undesirable Event
Undesirable Condition
«Classifies»
harms
0..*
Asset
Dangerous Event
has failed objectives
0..*
of
1..*
has condition
0..*
Vulnerability
allowed by
0..*
caused by
0..*
«Classifies»
arises from incident
1..*
Cause of incident
leads to
0..*
Incident
has
0..*
exploits
0..*
exploits
0..*
Exploit of Vulnerability
«Subsets»
exploited by fault
0..*
Fault
manifests as
1..*
«Subsets»
Physical Vulnerability
Cyber Vulnerability
result of
1..*
Failure
impacted by
0..*
Threat Actor
Undesirable Condition
exploited by
0..*
Danger Source
«Subsets»
Dangerous Condition
Dangerous Event
Unwitting Participant

For this article, it has only been possible to provide a brief introduction to the ontology with only a small subset of the model being presented. For further information and a much more detailed description of the model, refer to the paper "Formalising the Language of Risk" published as part of the 2020 Safety-Critical Systems Club Symposium Proceedings [4].

Conclusion

This article has set out to introduce the Risk Ontology that the Data Safety Initiative Working Group has assembled. The language is self-consistent (by virtue of the ontology formalism), and provides the means for describing causal situations that can result in harm to assets. The power of expression in the language is aided by the ontology classification meta-language as it allows terms to inherit higher order concepts.

An example being that whilst harm is a consequence that results from some other situation, harm is also a situation. This allows causal modelling to show, for example, how harms can propagate. For example, a fire in a bin (a localised harm) spreads (due to lack of adequate containment and/or proximity to other combustible materials) to destroy the building (a larger scale harm). Moreover, the causality modelling afforded by the situation related terms can be used in incident investigation (i.e. after the fact) where evidence is being assessed to determine why and how something occurred.

The ontology attempts to find common ground between the safety and security risk analysis by using unified terms such as "incident" and "vulnerability". The language described may not cover all aspects of safety or security analysis, but it is hoped that it provides enough common language to enable greater productivity in achieving security informed system safety.

Acknowledgements

The author would like to acknowledge the significant work of the "Threat and Risk Community" (threatrisk.org) in creating an ontology that paved the way for the creation of the DSIWG's own Risk Ontology. In that respect, our ontology shares many of the same concepts and terms as the Threat & Risk Ontology, although ours is smaller, in part because it lacks some of the foundational terms that the Threat & Risk Ontology formally defines.

The author would also like to thank the following people for their contribution to this paper and to the Risk Ontology: Paul Hampton, Divya Atkins, Martin Atkins, and Mike Parsons.

[1] "Risk Management - Guidelines," ISO 31000, 2018.

[2] "Risk Management - Vocabulary," ISO Guide 73, 2009.

[3] "Information technology - Object Management Group Unified Modeling Language (OMG UML), Superstructure," ISO/IEC 19505-2, 2012.

[4] "Assuring Safe Autonomy: Proceedings of the 28th Safety-Critical Systems Symposium (SSS'20) York, UK, 11th-13th February 2020", SCSC, January 2020, https://scsc.uk/SCSC-154

Dave Banham, Functional Safety Specialist at BlackBerry QNX

Dave Banham is a Chartered Engineer specialising in dependable software intensive and cyber physical systems. His professional career spans 28+ years including time at GEC, Alstom, Areva, Rolls-Royce, and most recently he joined the functional safety team at BlackBerry QNX. He is an active member of the DSIWG, MISRA C, MISRA C++ , and OMG SysML 1.7 working groups.