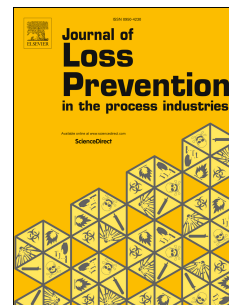


# Journal Pre-proof

Ontology-Driven Representation and Reasoning for Hazard and Operability Studies

Kateryna Miniailo, Zhouxiang Fei, Neil Buckley, Ogonnaya Anicho



PII: S0950-4230(26)00266-4

DOI: <https://doi.org/10.1016/j.jlp.2026.106177>

Reference: JLPP 106177

To appear in: *Journal of Loss Prevention in the Process Industries*

Received Date: 8 June 2026

Revised Date: 30 July 2026

Accepted Date: 25 August 2026

Please cite this article as: Miniailo, K., Fei, Z., Buckley, N., Anicho, O., Ontology-Driven Representation and Reasoning for Hazard and Operability Studies, *Journal of Loss Prevention in the Process Industries*, <https://doi.org/10.1016/j.jlp.2026.106177>.

This is a PDF of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability. This version will undergo additional copyediting, typesetting and review before it is published in its final form. As such, this version is no longer the Accepted Manuscript, but it is not yet the definitive Version of Record; we are providing this early version to give early visibility of the article. Please note that Elsevier's sharing policy for the Published Journal Article applies to this version, see: <https://www.elsevier.com/about/policies-and-standards/sharing#4-published-journal-article>. Please also note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2026 Published by Elsevier Ltd.

# Ontology-Driven Representation and Reasoning for Hazard and Operability Studies

Kateryna Miniailo<sup>a</sup>, Zhouxiang Fei<sup>a,1</sup>, Neil Buckley<sup>a</sup>, Ogbonnaya Anicho<sup>a</sup>

<sup>a</sup>*Liverpool Hope University, Taggart Avenue, Liverpool, L16 9JD, United Kingdom*

---

## Abstract

Within Process Safety Management, Process Hazard Analysis methods such as Hazard and Operability (HAZOP) studies provide a structured basis for identifying credible deviation scenarios and demonstrating that risks are understood and managed. In practice, however, HAZOP outputs are typically recorded as heterogeneous tables or narrative text, making reuse difficult and complicating systematic checks for missing causes, consequences, or safeguards. This paper presents an ontology-driven, reproducible pipeline that converts published chemical process HAZOP studies into a machine-readable knowledge graph, enabling structured analysis and completeness checking, while also creating a foundation for applying expert domain knowledge in retrieval-augmented generation and for future automation of HAZOP studies. An ontology in the Web Ontology Language (OWL-DL) was designed to capture core concepts from International Electrotechnical Commission standard IEC 61882:2016 for HAZOP studies, including node, guideword, deviation, cause, consequence, safeguard, and recommendation. The ontology was populated from a unified spreadsheet compiled from twelve published HAZOP studies using Protégé, an ontology editor, and its Cellfie plugin for spreadsheet-to-ontology transformation, while graph-based exploration was supported via Neo4j, a graph database platform. Logical consistency was verified using HermiT, an OWL reasoner, constituting the genuine description-logic reasoning step in this study, while a separate, non-inferential completeness audit – a property-presence check implemented in Python with Owlready2 – was used to evaluate whether each deviation instance was linked to the required core relationships. The ontology remained logically consistent, and the audit demonstrated that ontology-based completeness checking can reveal missing safety-relevant relationships in HAZOP records, supporting more systematic, traceable quality assurance and providing a practical step toward digitalised process safety workflows.

**Keywords:** HAZOP; knowledge modelling; ontology; process safety; semantic reasoning.

---

## 1. Introduction

The chemical and process industries encompass a wide range of operations, from petroleum refining and petrochemical manufacturing to pharmaceutical and food production, that routinely

---

<sup>1</sup> Corresponding author

handle flammable, toxic, or reactive materials at elevated temperatures and pressures. Process safety analysis is central to major accident prevention in the chemical and process industries, where hazardous inventories and operating conditions can escalate rapidly if deviations are not identified and controlled, as demonstrated by incidents such as the 2005 BP Texas City refinery explosion (U.S. Chemical Safety and Hazard Investigation Board, 2007; Basheer et al., 2019; Babaarslan et al., 2014). Within Process Safety Management (PSM), Process Hazard Analysis (PHA) provides the structured evidence base for identifying credible hazard scenarios and demonstrating that risks are understood and managed (Occupational Safety and Health Administration (OSHA), n.d.). Among PHA methods, the Hazard and Operability (HAZOP) study remains one of the most established and widely applied techniques for systematically identifying deviations from design intent, their causes, and their potential consequences (Kletz, 1999; Tripathi and Kushwaha, 2020).

IEC 61882:2016 was selected as the reference standard because it provides the current internationally recognised application guide for HAZOP studies, including the core procedural elements and documentation structure needed to define the ontology schema. According to this standard, a HAZOP study is performed by a multidisciplinary team following a defined workflow: the process is divided into nodes, guidewords are combined with process parameters to generate deviations, and each deviation is reviewed to document credible causes, consequences, existing safeguards, and recommendations for risk reduction, with clear traceability and reporting expectations (International Electrotechnical Commission (IEC), 2016). In practice, HAZOP teams typically examine process documentation (e.g., process flow diagrams and P&IDs) and capture results in tabular form to support review, sign-off, and subsequent design or operational changes (Mocellin et al., 2022). Despite its value and maturity, the traditional HAZOP process is time-consuming and labour-intensive, and it relies heavily on expert interpretation, which can introduce variability across teams and lead to inconsistent or incomplete coverage (Leone, 1996; Guimarães and Lapa, 2006). Missing or weakly documented entries in a HAZOP table can translate into overlooked hazard scenarios, untested assumptions about safeguards, and a false sense of completeness during safety assurance activities (Baybutt, 2015; Li and Zhao, 2025).

Computational approaches to supporting HAZOP have evolved considerably, from early rule-based expert systems through fuzzy-logic and hybrid risk-ranking methods to more recent ontology-based and natural language processing techniques that seek to make HAZOP knowledge machine-readable and reusable (Leone, 1996; Guimarães and Lapa, 2006; Batres and Kuraoka, 2003; Feng et al., 2021; Li and Zhao, 2024). Despite this progress, much of the knowledge captured in published HAZOP studies remains locked in unstructured narrative or semi-structured tables, limiting interoperability, systematic validity checking, and downstream reuse across studies and organisations (Li and Zhao, 2024; Single et al., 2019).

Closely related to the present work is the ontology-based framework of Single et al. (2020), which shares a similar ontology core and toolchain for representing HAZOP knowledge. That work, however, applies the ontology to actively infer and generate candidate deviation scenarios and is evaluated through quantitative benchmarking against a reference HAZOP, whereas the present study applies an ontology retrospectively to already-published HAZOP records, with the aim of auditing their completeness and internal consistency rather than generating new scenario content. This distinction between representing and quality-assuring existing documentation, on

the one hand, and generating new scenarios on the other, motivates the specific completeness-checking contribution developed in Section 4.

This study aims to contribute to the digitalisation of process safety by developing an ontology-driven approach for representing HAZOP knowledge in a machine-readable form. The ontology is intended as a domain-general schema for chemical-process HAZOP records, consistent with the sector-agnostic scope of IEC 61882 itself, rather than a model tailored to any single industry subsector. To that end, the research defines an ontology that captures core HAZOP concepts and relationships as specified in IEC 61882:2016, populates it from a unified dataset compiled from twelve published chemical-process HAZOP studies, and evaluates the resulting model through logical consistency and completeness checks. The completeness-checking methodology applied in this evaluation is adapted from the property-presence auditing approach of Li and Zhao (2025). Together, these steps support the broader goal of improving the structure, traceability, and reusability of HAZOP data.

Translating HAZOP knowledge into a form amenable to computational reasoning requires two preparatory stages: assembling a representative dataset from heterogeneous published sources, and developing an ontological framework that captures the conceptual structure of HAZOP analysis.

The remainder of this paper is organised as follows. Section 2 describes the construction of a unified dataset from twelve published HAZOP studies, covering data collection and preprocessing. Section 3 presents the ontology development process, including design, population, and graph-based visualisation. Together, these two sections establish the representational foundation on which the reasoning and completeness evaluation in Section 4 are built. Section 5 discusses implications and future directions, and Section 6 concludes.

## 2. Dataset Construction

### 2.1. Data Collection

Twelve published HAZOP case studies and HAZOP-style worksheets were collected from a mix of government guidance, peer-reviewed journals, and conference proceedings to build a diverse dataset for ontology construction. The corpus comprised one U.S. Department of Energy (DOE) technical handbook/report (n=1; (U.S. Department of Energy, 2004)), eight journal articles spanning process safety, environmental/bioprocess safety, nuclear engineering, occupational health, and chemical engineering application domains (n=8; (Wang et al., 2013; Severi et al., 2022; Guimarães and Lapa, 2006; Labovský et al., 2006; Habibi et al., 2008; Ora et al., 2017; Abbasi et al., 2018; Tripathi and Kushwaha, 2020)), and three proceedings-style publications reporting applied HAZOP analyses and extensions (n=3; IFAC-PapersOnLine n=2: (Arena et al., 2018; Mokhtarname et al., 2020); Chemical Engineering Transactions n=1: (Babaarslan et al., 2014)). The selected studies cover diverse systems and operating contexts — including chemical mixing, petrochemical refining/distillation, polymerisation reactors, and ethylene oxide/ethylene glycol production — capturing variation in node definition, guideword usage, and documentation detail, and thereby supporting a more robust ontology grounded in heterogeneous real-world HAZOP records. Studies were therefore deliberately drawn from multiple process domains so that the core schema, and in particular the Node and Deviation classes, would need to accommodate

heterogeneous operating contexts rather than a single sector. This diversity was a design choice for robustness rather than a dedicated cross-domain evaluation; the ontology's logical consistency across all populated domains (Section 4.1) offers indirect support that the schema held up under this heterogeneity, though no targeted generalisability test was conducted.

Together, these sources represent traditional, hybrid, and computationally enhanced HAZOP methodologies. Following the approach of Feng et al. (2021), unstructured text data from published reports were manually extracted into a structured spreadsheet format to facilitate subsequent ontology population.

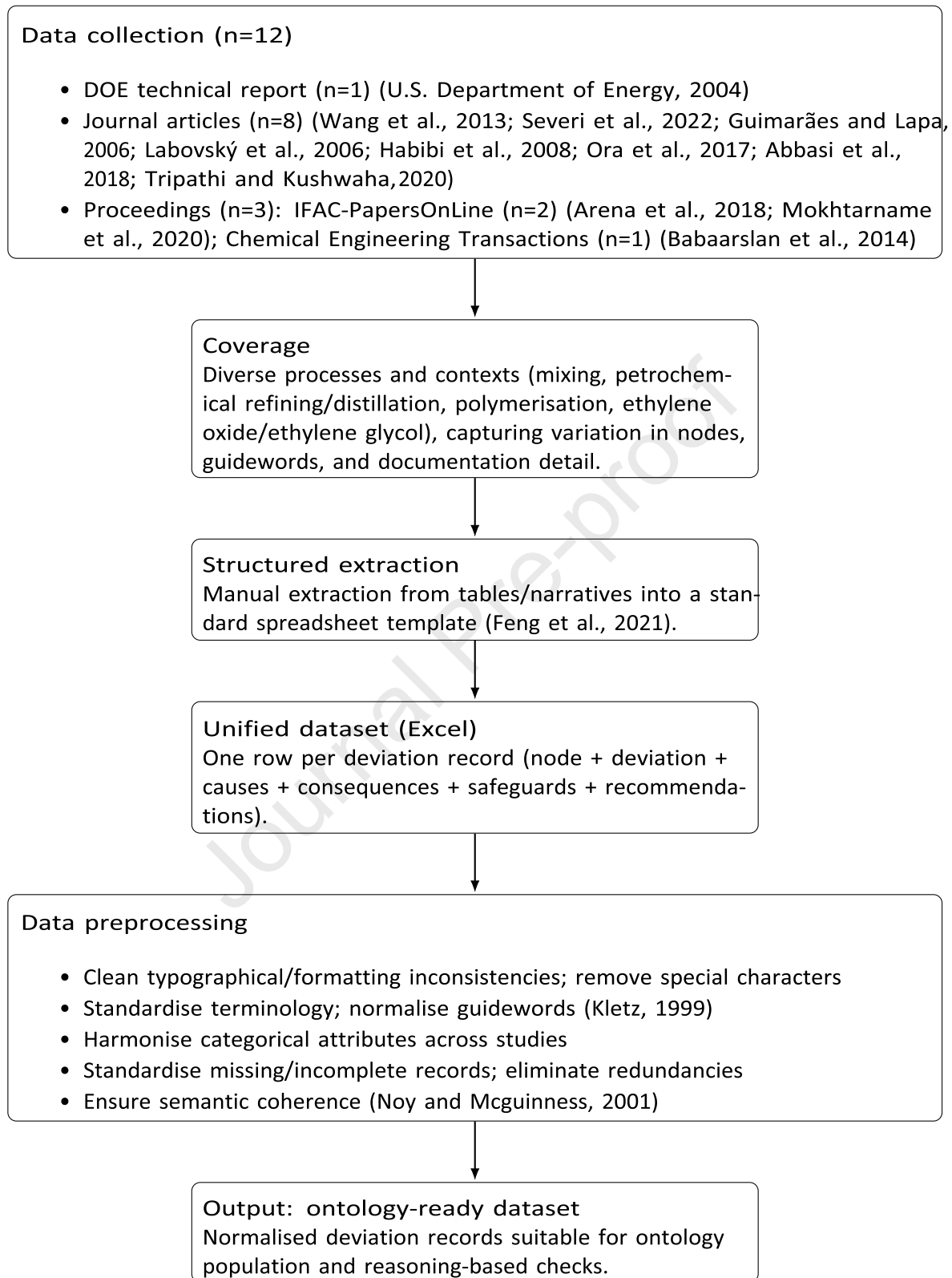


Figure 1: Workflow for data collection and preprocessing of published HAZOP studies prior to ontology population.

## 2.2. Data Preprocessing

The extracted HAZOP tables were compiled in Microsoft Excel, resulting in a unified dataset that integrated all twelve studies. Each entry corresponded to a unique HAZOP deviation record, typically consisting of the process node, parameter, guideword, deviation description, associated cause, consequence, safeguard, and recommendation. Data preprocessing involved cleaning typographical inconsistencies, standardising terminology (e.g., unifying guidewords such as *More*, *Less*, and *No*), harmonising categorical fields such as guidewords and parameters across all studies, and standardising missing or incomplete entries using explicit placeholders where necessary. Special characters and non-alphanumeric symbols were removed or normalised where required to ensure compatibility with the Cellfie spreadsheet-to-ontology transformation process.

This preprocessing stage ensured semantic coherence and eliminated redundancies, aligning with best practices for ontology development outlined by Noy and McGuinness (2001). This preprocessing produced a structured, machine-readable representation of the source material, in which HAZOP entries were encoded in a consistent format that software could automatically transform, query, and analyse, rather than leaving them as narrative or visually formatted tables intended only for human reading. Preprocessing was carried out manually using Microsoft Excel rather than a bespoke automated pipeline; Section 5 discusses NLP-based extraction tools as a direction for reducing the manual effort required at this stage in future work.

During preprocessing, source studies that systematically lacked entire relationship categories were excluded from the ontology-ready dataset to avoid distorting the completeness evaluation. Four of the twelve source studies were excluded at the population stage on this basis. The bioprocess subsystem HAZOP from Severi et al. (2022), comprising 179 deviation records, documented causes and consequences but recorded no safeguards or recommendations for any entry. The polymerisation and ethylene oxide reactor records from Labovský et al. (2006) (6 records) omitted all safeguard information. The esterifier and melter unit records from Abbasi et al. (2018) (4 records) likewise recorded no safeguards or recommendations for any entry. The cooling-water chlorination worked example from U.S. Department of Energy (2004) (5 records) omitted a recommendation for every entry and a consequence or safeguard for the majority of entries. Including these studies would have inflated the proportion of missing safeguards, recommendations, or consequences without reflecting genuine variation in documentation practice across studies. Individual rows from the remaining studies were also excluded where the source HAZOP team had identified no credible deviation for a given guideword combination (e.g., entries marked "N/A" with no associated causes, consequences, or safeguards). After these exclusions, the ontology-ready dataset comprised 58 deviation records drawn from eight of the twelve source studies. All twelve studies contributed to the overall data collection and informed the ontology design; the exclusions applied only to the population stage.

| ID | Node                              | Parameter     | Guide-word | Deviation                           | Cause                                                                     | Consequence                               | Safeguard                              | Recommendation                                            |
|----|-----------------------------------|---------------|------------|-------------------------------------|---------------------------------------------------------------------------|-------------------------------------------|----------------------------------------|-----------------------------------------------------------|
| E1 | Batch mixer – Acetic-acid feed    | Flow          | More       | High flow                           | Valve malfunction                                                         | Overacidic mixture; corrosion risk        | —                                      | Inspect valve; fit high-level alarm                       |
| E2 | Batch mixer – 33 percent HCl feed | Flow          | More       | High flow                           | Acid passes through valve when closed                                     | Mixer overflow; corrosive mixture         | —                                      | Regulate throttle; add high-level trip                    |
| E3 | Raw-water pump and header         | Flow          | More       | High flow                           | Accidental start of a second pump; manual valve wide open; flange failure | Line trip; pipe cracking                  | PLC logic prevents starting extra pump | Program PLC to inhibit second-pump start; inspect flanges |
| E4 | Acid transfer                     | Concentration | Less       | Acid concentration below 98 percent | Water or iron impurities; moisture ingress                                | Equipment corrosion; reduced acid potency | Periodic analysis                      | Humidity trap in air line; routine sampling               |
| E5 | Acid transfer                     | Concentration | More       | Acid concentration above 4 percent  | Overspeed acid pump; low water flow in entry of mixing tee                | Resin damage; equipment corrosion         | Flow-switch high trip                  | Increase acid injection pump course                       |

Table 1: Excerpt of the unified HAZOP dataset showing example deviation records used for ontology population.

For clarity in the subsequent sections, the five illustrative records are referenced as Examples E1–E5, corresponding respectively to Batch mixer – Acetic-acid feed (High flow), Batch mixer – 33 percent HCl feed (High flow), Raw-water pump and header (High flow), Acid transfer (Acid concentration below 98 percent), and Acid transfer (Acid concentration above 4 percent).

### 3. Ontology Development

#### 3.1. Ontology Design

The ontology was developed in the Web Ontology Language (OWL) using Protégé, following the guidelines from *Ontology Development 101* (Noy and McGuinness, 2001). Its conceptual

structure drew inspiration from earlier ontological representations of HAZOP information, notably Batres and Kuraoka (2003), who modelled HAZOP concepts through classes such as *Deviation*, *Cause*, and *Consequence*, and the OntoCAPE process ontology (Morbach et al., 2007), which provides a hierarchical modular framework for process engineering knowledge.

The top-level classes included *Node*, *Deviation*, *Cause*, *Consequence*, *Guideword*, *Safeguard*, and *Recommendation*. Object properties were defined to represent the relationships among these entities, such as *hasCause*, *hasConsequence*, *hasGuideword*, *hasSafeguard*, and *hasRecommendation*. These relations capture the causal and logical dependencies inherent in HAZOP studies, formally expressing knowledge such as: “A deviation has one or more causes, which may lead to one or more consequences.”

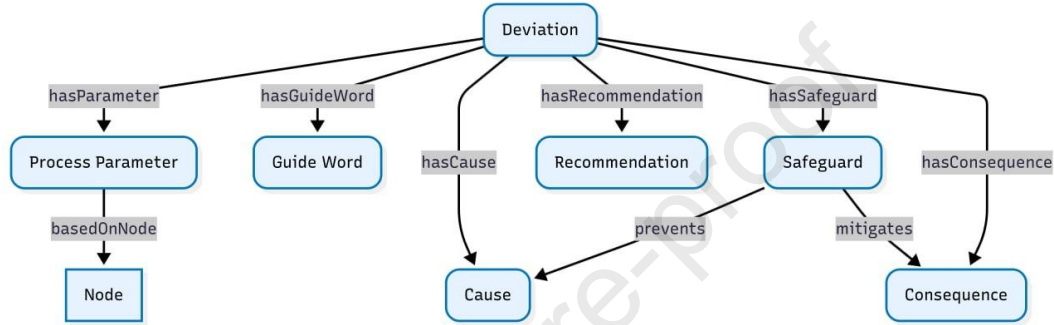


Figure 2: Class hierarchy and relationships within the developed HAZOP ontology.

Each class was annotated with definitions and examples drawn from the dataset to ensure semantic clarity. The ontology adhered to the OWL-DL profile to maintain reasoning compatibility while providing sufficient expressiveness for inference-based tasks.

### 3.2. Ontology Population

The populated ontology was generated using the Protégé Cellfie plugin, which facilitates the transformation of structured tabular data into ontology individuals and property assertions (Tudorache et al., 2013; Stanford Center for Biomedical Informatics Research, n.d.). Mapping rules were defined for each column in the spreadsheet: for instance, each deviation was instantiated as an individual of the class *Deviation*, linked to corresponding *Cause*, *Consequence*, and *Safeguard* individuals via the defined object properties. The mapping followed a consistent pattern:

- "Deviation" column → individuals of class *Deviation*
- "Cause" column → individuals of class *Cause*
- "Consequence" column → individuals of class *Consequence*
- "Safeguard" column → individuals of class *Safeguard*
- "Recommendation" column → individuals of class *Recommendation*

Transformation rules were iteratively validated to ensure accurate instantiation of relationships, reducing mismatches or null references. Following population, the ontology comprised the class, object property, individual, and axiom counts summarised in Table 2, effectively transforming the raw HAZOP tabular data into a structured, queryable knowledge graph.

Table 2: Ontology size metrics after population, reported for comparability with prior HAZOP ontologies.

| Metric            | Count |
|-------------------|-------|
| Classes           | 8     |
| Object properties | 9     |
| Individuals       | 296   |
| Axioms (total)    | 1306  |
| Logical axioms    | 697   |

Because the Cellfie transformation instantiates one *Deviation* individual per unique deviation description, multiple spreadsheet rows sharing the same label (e.g., *High flow* occurring across three different process nodes) contribute their relationships to a single ontology entity. The 58 source rows yielded 46 unique *Deviation* individuals; a relationship such as *hasSafeguard* is considered present for a given deviation if any contributing row supplies it.

### 3.3. Visualisation of Ontology Relationships

The populated ontology was exported into a graph format and visualised using Neo4j, an open-source graph database suited for exploring semantic relationships. In this representation, nodes corresponded to ontology individuals (e.g., deviations or causes), and edges represented semantic relationships such as *hasCause* or *hasSafeguard*. Graph visualisation supported exploratory analysis and validation by revealing frequently recurring deviations (e.g., *High Pressure*, *Low Flow*) and their typical causes or safeguards. This visual inspection phase contributed to verifying ontology completeness and internal consistency before reasoning and inference experiments.

Figure 3 illustrates how tabular HAZOP entries can be transformed into ontology relationships. The subgraph is grounded in the three *High flow* examples shown in Table 1, namely E1 (Batch mixer – Acetic-acid feed), E2 (Batch mixer – 33 percent HCl feed), and E3 (Raw-water pump and header). The central node denotes the shared deviation concept, while the surrounding nodes represent associated causes and consequences extracted from these records. In particular, the relationship between *Valve malfunction* and *High flow*, as well as the consequence *Over-acidic mixture*, are based on Table 2 of Ora et al. (2017); the relationship involving *Acid passes through valve when closed* is based on Table 3 of Ora et al. (2017); and the cause *Accidental start of a second pump* is drawn from Table 3 of Habibi et al. (2008). This example shows how multiple tabular records that share the same deviation can be integrated into a connected graph structure while preserving the provenance of their individual safety-relevant relationships.

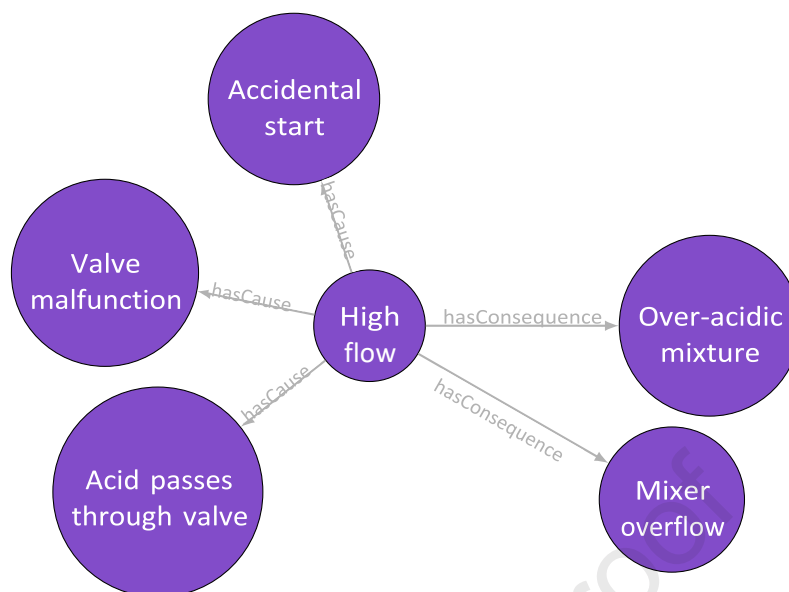


Figure 3: Illustrative ontology subgraph for the deviation *High flow*, derived from examples E1–E3 in Table 1. The figure combines relationships reported across multiple source records: *Valve malfunction* and *Overacidic mixture* from Table 2 of Ora et al. (2017), *Acid passes through valve when closed* from Table 3 of Ora et al. (2017), and *Accidental start of a second pump* from Table 3 of Habibi et al. (2008). Node labels are abbreviated for readability. The three *hasCause* edges shown originate from three distinct process nodes (Batch mixer – Acetic-acid feed, Batch mixer – 33 percent HCl feed, and Raw-water pump and header) that happen to share the *High flow* deviation label; they do not represent three causes co-occurring within a single shared context.

Figure 4 illustrates how the ontology can represent less frequently occurring, process-specific deviations within the same semantic framework used for more common HAZOP patterns. In contrast to the *High flow* deviation shown in Figure 3, which appears repeatedly across several records and source studies, the *Acid concentration* deviations are comparatively rare and are tied specifically to the *Acid transfer* node derived from Table 4 of Habibi et al. (2008). The figure is based on the two corresponding records represented in Table 1, namely the deviations *Acid concentration below 98%* and *Acid concentration above 4%*. By linking both deviation instances to the shared process node *Acid transfer* through the relationship *basedOnNode*, the ontology makes their common operational context explicit while preserving selected guideword and causal relationships for both cases, together with additional safeguard, consequence, and recommendation links for the *Acid concentration above 4%* deviation. Although these deviations are less common than *High flow*, they are still connected to broader patterns in the dataset through the shared guidewords *Less* and *More*, which also occur in other HAZOP records involving different parameters and process nodes.

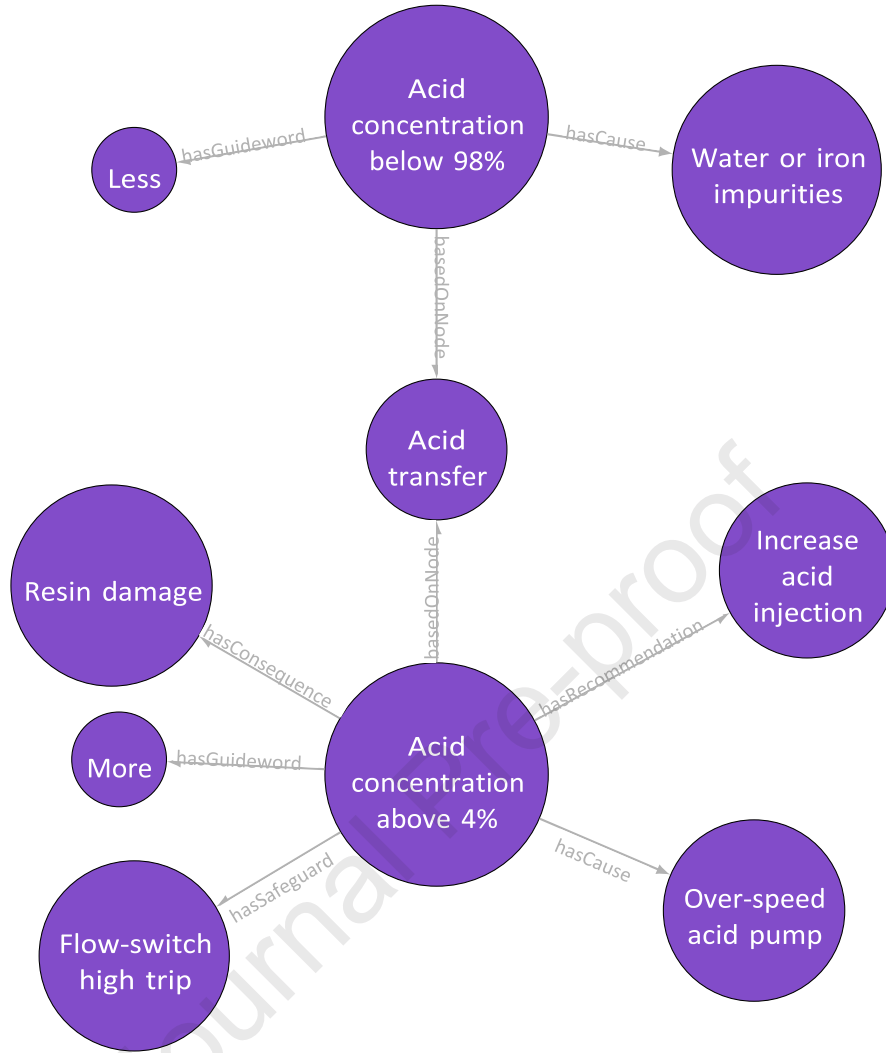


Figure 4: Ontology subgraph derived from the *Acid transfer* deviations reported in Table 4 of Habibi et al. (2008). The figure shows two comparatively rare deviation instances, *Acid concentration below 98%* and *Acid concentration above 4%*, linked to the shared process node *Acid transfer* through *basedOnNode*. Additional nodes illustrate selected guideword, cause, safeguard, consequence, and recommendation relationships associated with the *Acid concentration above 4%* deviation.

#### 4. Reasoning Experiment and Completeness Evaluation

The dataset underlying the evaluations in this section is the 58-record ontology-ready set described in Section 2.2, following the exclusion of four source studies whose systematic lack of safeguard or recommendation information would otherwise have distorted the completeness estimates reported below.

##### 4.1. Consistency Checking

To verify the logical soundness of the developed ontology, a reasoning experiment was conducted in Protégé using the Hermit reasoner (Shearer et al., 2008), an OWL description logic reasoner that supports automated inference and consistency checking. The aim was to ensure

that all class definitions, object properties, and individual assertions were logically consistent. The reasoning process successfully completed in 184 ms, confirming that no unsatisfiable classes or conflicting restrictions were present. The reasoner also inferred class and object property hierarchies, validating that the ontology adhered to the OWL DL profile and was suitable for automated reasoning tasks.

To illustrate what a consistency violation would look like in this ontology: if, for example, *Safeguard* and *Cause* were declared disjoint classes – a reasonable modelling choice, since a protective measure and a causal factor are conceptually distinct – then an extraction or population error that asserted a single individual to be an instance of both classes (e.g., a row miscoded so that a safeguard description was linked via *hasCause* instead of *hasSafeguard*) would cause Hermit to classify that individual, and the ontology as a whole, as inconsistent.

#### 4.2. Completeness Checking

Beyond logical consistency, the ontology was evaluated for semantic completeness through a set of automated checks implemented in Python using the owlready2 library. This completeness check is a property-presence audit rather than a description-logic inference: it does not derive new facts about a deviation but tests, for each existing individual, whether an expected relationship was asserted during population. Any relationship reported as missing therefore reflects an absence in the original source record or its manual extraction, not a gap identified through logical inference. Following the approach proposed by Li and Zhao (2025), the completeness check assessed whether each *Deviation* individual possessed the core relationships required for a fully defined HAZOP entry. Four object properties were evaluated: *hasGuideword*, *hasSafeguard*, *hasConsequence*, and *hasCause*. A relation was considered missing when the associated instance was absent or labelled only with a placeholder (e.g., “—”), reflecting an incomplete or undefined entry in the source dataset. Because multiple source rows sharing the same deviation label are merged into a single *Deviation* individual during ontology population (Section 3.2), the completeness check operates over unique deviations rather than raw spreadsheet rows: a relationship is counted as present for a deviation if at least one contributing row supplies it. This is the process by which the 58 ontology-ready records reduce to the 46 *Deviation* individuals evaluated below.

The completeness script automatically loaded the ontology, traversed all instances of the *Deviation* class, and produced a tabulated summary of missing or incomplete relationships. The results are summarised in Table 3.

Table 3: Completeness check results for Deviation relationships. Total = 46 refers to the number of unique *Deviation* individuals in the populated ontology after merging rows by shared deviation label (Section 3.2).

| Check                         | Missing | Total |
|-------------------------------|---------|-------|
| Deviation_without_Guideword   | 0       | 46    |
| Deviation_without_Safeguard   | 13      |       |
| Deviation_without_Consequence | 2       |       |
| Deviation_without_Cause       | 1       |       |

### 4.3. Results and Interpretation

The reasoning experiment confirmed that the ontology was logically consistent and sufficiently expressive to represent the core relationships required for HAZOP knowledge modelling. The completeness evaluation showed that guideword assignments were fully preserved across the dataset, with all 46 deviations linked to valid guidewords. In contrast, incompleteness was concentrated mainly in safeguard information: 13 of 46 deviations (28.3%) lacked an associated safeguard. Because the merge aggregates relationships across all source rows for a given deviation label, this proportion represents a lower bound; at the individual record level, 19 of the 58 source rows (32.8%) lacked safeguard information. Missing consequences were much less frequent, affecting 2 deviations (4.3%), while only 1 deviation (2.2%) lacked a cause.

At a higher level, these results suggest that the published HAZOP records used in the dataset were generally strong in capturing the basic deviation structure, particularly the guideword–deviation relationship, but less consistent in documenting downstream protective and explanatory information. In other words, the core identification of deviations was usually present, whereas the supporting safety context was more unevenly recorded. This pattern is also visible in the illustrative excerpt in Table 1, where examples E1 and E2 contain missing safeguard entries, represented by placeholders and therefore identified by the automated check as incomplete. Such omissions are important because missing safeguards, causes, or consequences can reduce the reliability and reusability of HAZOP documentation in later review, audit, or decision-support settings.

These gaps appear to stem primarily from incompleteness or ambiguity in the original HAZOP tables extracted from the literature rather than from structural limitations of the ontology itself. The ability to detect such patterns automatically demonstrates the practical value of the ontology for supporting traceable quality assurance and validation of HAZOP datasets.

## 5. Discussion

The central contribution of this work is the unified, multi-study HAZOP knowledge graph itself – assembled from published studies into a single, queryable representation – rather than the ontology schema in isolation, which follows largely established modelling practice (Batres and Kuraoka (2003); Morbach et al. (2007)). Beyond the consistency and completeness analyses reported here, this aggregated resource is intended as a reusable seed corpus or benchmark for future NLP- and retrieval-augmented-generation (RAG) based HAZOP research.

This work demonstrates the feasibility of representing HAZOP knowledge through ontology-driven models, highlighting how semantic reasoning can improve data quality, traceability, and reuse. Ontology-based frameworks provide a structured foundation for integrating heterogeneous safety information and verifying its internal coherence, complementing traditional manual analyses (Batres and Kuraoka, 2003; Morbach et al., 2007). Reasoning tools such as HerMiT (Shearer et al., 2008) and Python libraries like owlready2 (Lamy, 2017) enable automated validation of relationships between deviations, causes, and safeguards, supporting more systematic and reproducible hazard assessments.

A key insight from this research is that the completeness and accuracy of ontology-based reasoning depend strongly on the quality and structure of source data. As noted by Dunj6 et al.

(2010), HAZOP documentation across the chemical industry remains highly variable, often lacking consistent terminology or standardised tabular formats. Such heterogeneity poses challenges for data extraction and necessitates significant manual preprocessing. Addressing this issue requires greater standardisation in how HAZOP results are reported and, ideally, the development of digital formats that facilitate automatic ingestion into knowledge-based systems.

This domain-general approach also has limitations. Node definitions are not always directly comparable across studies: the granularity at which a "node" is defined (e.g., a single vessel versus an entire process section) varies with each source study's own scope and reporting conventions, which may affect cross-study aggregation at the node level even though deviation-level completeness checking remains valid. Relatedly, the manual extraction of HAZOP records from twelve heterogeneous published sources was not independently validated (e.g., via a formal inter-rater agreement check); future work should incorporate such validation to strengthen confidence in the fidelity of the extracted dataset. A further limitation concerns the merging of deviations that share the same label across unrelated process nodes (Section 3.2, Figure 3): because the completeness check operates on merged *Deviation* individuals, a deviation with contributing rows from several nodes can appear more complete than any single underlying record, and the graph visualisation can visually imply an association between causes or consequences drawn from different, unrelated nodes even though no single real deviation instance exhibits all of them together. Finally, the 46–58 record sample underlying the headline completeness figures is small and was not randomly selected – it comprises whichever published HAZOP studies were accessible and sufficiently documented for extraction – so the reported percentages (e.g., the 28.3% missing-safeguard rate) should be read as descriptive of this particular corpus rather than as generalisable estimates of documentation quality across HAZOP practice more broadly. This risk is compounded by the exclusion criteria applied during population (Section 2.2): because studies that systematically lacked safeguards or recommendations were removed prior to the completeness evaluation, the retained sample is biased toward better-documented studies, meaning the reported missing-safeguard rate likely understates the incompleteness that would be observed across the full, unfiltered set of twelve source studies.

Recent advances in natural language processing (NLP) have opened promising pathways for automating the acquisition of safety information. Studies have shown that NLP techniques can identify entities, causal relations, and guideword–parameter patterns in unstructured HAZOP reports or incident descriptions (Feng et al., 2021; Li and Zhao, 2024; Elhosary and Moselhi, 2025). Integrating these methods with ontological reasoning could enable automated population of knowledge graphs directly from textual data, significantly reducing the manual effort required for data preparation. Furthermore, large language models (LLMs) trained on domain-specific corpora could support semi-automated HAZOP generation by proposing candidate causes, consequences, and safeguards that are subsequently validated against ontology-defined logical constraints.

Automating data acquisition and integration also aligns with the broader trend toward digital process safety management. Coupling ontology-based reasoning with intelligent data pipelines would allow continuous ingestion of operational data, incident reports, and maintenance logs, providing real-time feedback on process integrity. This vision echoes recent efforts to bridge knowledge representation and industrial data analytics in chemical engineering (Wang et al.,

2022). Such integration could transform HAZOP from a static documentation process into a dynamic, data-driven decision-support system.

Future work should focus on extending this ontology with probabilistic and temporal reasoning capabilities to capture dependencies among process parameters and evolving system states. Expanding interoperability with established engineering ontologies such as OntoCAPE, a modular ontology framework for process systems engineering and chemical plant knowledge representation (Morbach et al., 2007), would facilitate seamless integration into larger process modelling environments, advancing the digitalisation of hazard analysis. At larger scale, this completeness-audited, consistency-verified knowledge graph may also serve as a complementary foundation for prospective scenario-generation methods such as that of Single et al. (2020), though this complementarity remains to be empirically tested.

## 6. Conclusion

This study has shown that HAZOP knowledge can be formalised in an ontology-driven, machine-readable representation that preserves the core semantic structure of deviation analysis, including relationships among nodes, guidewords, causes, consequences, safeguards, and recommendations. By converting heterogeneous published HAZOP records into a unified knowledge graph, the work demonstrates a practical approach to making process safety information more structured, traceable, and reusable.

The results indicate that ontology-based reasoning can support not only logical validation of the resulting representation but also systematic assessment of dataset completeness. In particular, the study showed that while core deviation structures were generally well captured in the source material, supporting safety information such as safeguards was less consistently documented. This highlights the value of semantic modelling as a means of exposing incompleteness in published HAZOP records and strengthening the quality assurance of safety-critical information.

More broadly, this work contributes to the ongoing digitalisation of process safety by showing how formal knowledge representation can enhance the consistency, transparency, and interoperability of HAZOP data. The proposed approach provides a foundation for future developments in automated knowledge extraction, integration with wider engineering knowledge frameworks, and intelligent decision-support systems for hazard analysis.

## Funding

This project received funding of Undergraduate Summer Research Scholarship from the Pro-Vice-Chancellor (PVC) at Liverpool Hope University.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## CRediT authorship contribution statement

Kateryna Minaiilo: Conceptualisation, Methodology, Software, Data curation, Formal analysis, Investigation, Visualisation, Writing — original draft, Writing — review & editing. Zhouxiang Fei: Supervision, Methodology, Writing — review & editing. Neil Buckley: Writing — review & editing. Ogbonnaya Anicho: Writing — review & editing.

### Data availability

Data will be made available on request.

### Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used ChatGPT (OpenAI) and Claude (Anthropic) to improve clarity and readability, assist with drafting and rephrasing sections, and refine grammar and academic style. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

### References

Abbasi, T., Abbasi, T., Abbasi, S.A., 2018. Role of hazop in assessing risk of accidents in chemical process industries: Capability and lacunae. *International Journal of Engineering, Science and Technology*.

Arena, D., Criscione, F., Trapani, N., 2018. Risk assessment in a chemical plant with a cpn hazop tool. *IFAC-PapersOnLine* 51, 939–944. URL: <https://www.sciencedirect.com/science/article/pii/S2405896318316173>, doi: <https://doi.org/10.1016/j.ifacol.2018.08.487>. 16th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2018.

Babaarslan, N., Alpbaz, M., Cabbar, Y., 2014. Accidental risk assessment on atmospheric distillation column in oil refinery. *Chemical Engineering Transactions* 36, 475–480. URL: <https://www.cetjournal.it/index.php/cet/article/view/CET1436080>, doi: [10.3303/CET1436080](https://doi.org/10.3303/CET1436080).

Basheer, A., Tauseef, S., Abbasi, T., Abbasi, S.A., 2019. Methodologies for assessing risks of accidents in chemical process industries. *Journal of Failure Analysis and Prevention* 19, 623–648. URL: <https://doi.org/10.1007/s11668-019-00642-w>, doi: [10.1007/s11668-019-00642-w](https://doi.org/10.1007/s11668-019-00642-w).

Batres, R., Kuraoka, K., 2003. An ontological approach to represent HAZOP information. Technical Report TR-2003-01. Process Systems Engineering Laboratory, Tokyo Institute of Technology. doi: [10.13140/RG.2.2.27761.56165](https://doi.org/10.13140/RG.2.2.27761.56165).

Baybutt, P., 2015. A critique of the hazard and operability (hazop) study. *Journal of Loss Prevention in the Process Industries* 33, 52–58. URL:

<https://www.sciencedirect.com/science/article/pii/S0950423014001983>, doi:<https://doi.org/10.1016/j.ilp.2014.11.010>.

Dunjó, J., Fthenakis, V.M., Vílchez, J.A., Arnaldos, J., 2010. Hazard and operability (hazop) analysis. a literature review. *Journal of hazardous materials* 173 1-3, 19–32. URL: <https://api.semanticscholar.org/CorpusID:23643048>.

Elhosary, E., Moselhi, O., 2025. Evaluating natural language processing algorithms for improved hazard and operability analysis. *Geodata and AI* 4, 100026. URL: <https://www.sciencedirect.com/science/article/pii/S3050483X25000255>, doi:<https://doi.org/10.1016/j.geoai.2025.100026>.

Feng, X., Dai, Y., Ji, X., Zhou, L., Dang, Y., 2021. Application of natural language processing in hazop reports. *Process Safety and Environmental Protection* 155, 41–48. URL: <https://www.sciencedirect.com/science/article/pii/S0957582021004675>, doi:<https://doi.org/10.1016/j.psep.2021.09.001>.

Guimarães, A.C.F., Lapa, C.M.F., 2006. Hazard and operability study using approximate reasoning in light-water reactors passive systems. *Nuclear Engineering and Design* 236, 1256–1263. URL: <https://www.sciencedirect.com/science/article/pii/S0029549305004231>, doi:[doi:10.1016/j.nucengdes.2005.11.005](https://doi.org/10.1016/j.nucengdes.2005.11.005).

Habibi, E., Zare, M., Barkhordari, A., Mirmohammadi, S., Halvani, G., 2008. Application of a hazard and operability study method to hazard evaluation of a chemical unit of the power station. *J Res Health Sci* 8, 13–20. URL: <https://pubmed.ncbi.nlm.nih.gov/23344069/>, arXiv:<https://jrhs.umsha.ac.ir/PDF/263>.

International Electrotechnical Commission (IEC), 2016. IEC 61882:2016 Hazard and operability studies (HAZOP studies) – Application guide. International Standard IEC 61882:2016. International Electrotechnical Commission (IEC). Geneva, Switzerland. URL:<https://webstore.iec.ch/en/publication/24321>.

Kletz, T., 1999. HAZOP and HAZAN: Identifying and Assessing Process Industry Hazards. Institution of Chemical Engineers, Rugby, UK.

Labovský, J., Jelemenský, v., Markoš, J., 2006. Safety analysis and risk identification for a tubular reactor using the hazop methodology. *Chemical Papers* 60, 454–459. doi:[10.2478/s11696-006-0082-0](https://doi.org/10.2478/s11696-006-0082-0).

Lamy, J.B., 2017. Owlready: Ontology-oriented programming in python with automatic classification and high level constructs for biomedical ontologies. *Artificial Intelligence in*

Medicine 80, 11–28. URL: <https://www.sciencedirect.com/science/article/pii/S0933365717300271>, doi:<https://doi.org/10.1016/j.artmed.2017.07.002>.

Leone, H., 1996. A knowledge-based system for hazop studies.: The knowledge representation structure. *Computers & Chemical Engineering* 20, S369–S374. URL: <https://www.sciencedirect.com/science/article/pii/0098135496000725>, doi:[https://doi.org/10.1016/0098-1354\(96\)00072-5](https://doi.org/10.1016/0098-1354(96)00072-5). European Symposium on Computer Aided Process Engineering-6.

Li, Z., Zhao, J., 2024. Constructing a knowledge graph for automated hazop analysis, in: Manenti, F., Reklaitis, G.V. (Eds.), 34th European Symposium on Computer Aided Process Engineering / 15th International Symposium on Process Systems Engineering. Elsevier. volume 53 of *Computer Aided Chemical Engineering*, pp. 1291–1296. URL: <https://www.sciencedirect.com/science/article/pii/B9780443288241502167>, doi:<https://doi.org/10.1016/B978-0-443-28824-1.50216-7>.

Li, Z., Zhao, J., 2025. Hazopct: A hazop analysis completeness tool based on knowledge graph reasoning. *Process Safety and Environmental Protection* 197, 107025. URL: <https://www.sciencedirect.com/science/article/pii/S0957582025002927>, doi:<https://doi.org/10.1016/j.psep.2025.107025>.

Mocellin, P., De Tommaso, J., Vianello, C., Maschio, G., Saulnier-Bellemare, T., Virla, L.D., Patience, G.S., 2022. Experimental methods in chemical engineering: Hazard and operability analysis—hazop. *The Canadian Journal of Chemical Engineering* 100, 3450–3469. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/cjce.24520>, doi:<https://doi.org/10.1002/cjce.24520>.

Mokhtarname, R., Safavi, A., Urbas, L., Salimi, F., Zerafat, M., Harasi, N., 2020. Toward hazop 4.0 approach for managing the complexities of the hazard and operability of an industrial polymerization reactor, pp. 13593–13600. URL: <https://www.sciencedirect.com/science/article/pii/S2405896320311812>, doi: <https://doi.org/10.1016/j.ifacol.2020.12.852>. 21st IFAC World Congress.

Morbach, J., Yang, A., Marquardt, W., 2007. Ontocape-a large-scale ontology for chemical process engineering. *Engineering Applications of Artificial Intelligence* 20, 147–161. doi:[10.1016/j.engappai.2006.06.010](https://doi.org/10.1016/j.engappai.2006.06.010).

Noy, N., McGuinness, D., 2001. Ontology development 101: A guide to creating your first ontology. *Knowledge Systems Laboratory* 32.

Occupational Safety and Health Administration (OSHA), n.d. 29 cfr 1910.119 — process safety management of highly hazardous chemicals. Electronic Code of Federal Regulations (eCFR). URL: <https://www.ecfr.gov/current/title-29/subtitle-B/chapter-XVII/part-1910/subpart-H/section-1910.119>. paragraph (e) Process hazard analysis.

Ora, A., Nandan, A., Kumar, A., 2017. Hazop analysis of a chemical process plant: A case study. *International Journal of Engineering and Advanced Technology* 2, 79–84. URL: <https://www.ijarnd.com/manuscript/hazard-identification-chemical-mixing-plant-hazop-study/>.

Severi, C.A., Pérez, V., Pascual, C., Muñoz, R., Lebrero, R., 2022. Identification of critical operational hazards in a biogas upgrading pilot plant through a multi-criteria decision-making and ftopsis-hazop approach. *Chemosphere* 307, 135845. URL: <https://www.sciencedirect.com/science/article/pii/S0045653522023384>, doi:<https://doi.org/10.1016/j.chemosphere.2022.135845>.

Shearer, R., Motik, B., Horrocks, I., 2008. Hermit: A highly-efficient owl reasoner, in: *OWL: Experiences and Directions*. URL: <https://api.semanticscholar.org/CorpusID:7951194>.

Single, J.I., Schmidt, J., Denecke, J., 2019. Computer-aided hazop studies: Knowledge representation and algorithmic hazard identification, in: *Safety and Security Engineering VIII*, WIT Press. pp. 55–66. doi:[10.2495/SAFE190061](https://doi.org/10.2495/SAFE190061).

Single, J.I., Schmidt, J., Denecke, J., 2020. Ontology-based computer aid for the automation of hazop studies. *Journal of Loss Prevention in the Process Industries* 68, 104321. URL: <https://www.sciencedirect.com/science/article/pii/S0950423020306082>, doi:<https://doi.org/10.1016/j.jlp.2020.104321>.

Stanford Center for Biomedical Informatics Research, n.d. Protégé cellfie: Mappingmaster and excel import plugin. <https://github.com/protegeproject/cellfie-plugin>.

Tripathi, K., Kushwaha, N., 2020. Hazard operability study for chemical industry. *International Journal for Scientific Research and Development* 8.

Tudorache, T., Nyulas, C., Noy, N.F., Musen, M.A., 2013. Webprotégé: A collaborative ontology editor and knowledge acquisition tool for the web, IOS Press, NLD. p. 89–99.

U.S. Chemical Safety and Hazard Investigation Board, 2007. Investigation Report: Refinery Explosion and Fire (15 Killed, 180 Injured). Technical Report 2005-04-I-TX. U.S. Chemical Safety and Hazard Investigation Board. Washington, DC. BP Texas City refinery, March 23, 2005.

U.S. Department of Energy, 2004. Chemical Process Hazards Analysis (DOE-HDBK1100-2004). Technical Report. Office of Environment, Safety and Health, Washington, DC. URL: <https://www.standards.doe.gov/standards-documents/1100/1100-BHdbk-2004/>.

Wang, F., Zhao, Y., Yang, O., Cai, J., Deng, M., 2013. Process safety data management program based on hazop analysis and its application to an ethylene oxide/ethylene glycol plant. *Journal of Loss Prevention in the Process Industries* 26, 1399–1406. URL: <https://www.sciencedirect.com/science/article/pii/S0950423013001757>, doi:<https://doi.org/10.1016/j.jlp.2013.08.020>.

Wang, Z., Zhang, B., Gao, D., 2022. A novel knowledge graph development for industry design: A case study on indirect coal liquefaction process. *Computers in Industry* 139, 103647. URL: <https://www.sciencedirect.com/science/article/pii/S0166361522000446>, doi:<https://doi.org/10.1016/j.compind.2022.103647>.

## Highlights (Reference of Article: JLPP 106177)

- OWL ontology captures core HAZOP concepts from IEC 61882:2016
- Twelve published HAZOP studies unified into one knowledge graph
- 28% of deviations lack safeguard links in published records
- Automated completeness checking via Python and Owlready2
- Reproducible pipeline from spreadsheet to queryable HAZOP graph

**Declaration of interests**

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: