Proceedings of the 33rd European Safety and Reliability Conference (ESREL 2023)
Edited by Mário P. Brito, Terje Aven, Piero Baraldi, Marko Čepin and Enrico Zio

Published by Research Publishing, Singapore.
doi: 10.3850/978-981-18-8071-1_P021-cd

Analyzing Hydrogen-Related Undesired Events: A Systematic Database for Safety Assessment

Alessandro Campari

Department of Mechanical and Industrial Engineering, NTNU, Norway. E-mail: alessandro.campari@ntnu.no

Elena Stefana

Department of Mechanical and Aerospace Engineering, Sapienza University of Rome, Italy.
E-mail: elena.stefana@uniroma1.it

Diletta Ferrazzano

Department of Civil, Chemical, Environmental and Materials Engineering, University of Bologna, Italy.
E-mail: diletta.ferrazzano@studio.unibo.it

Nicola Paltrinieri

Department of Mechanical and Industrial Engineering, NTNU, Norway. E-mail: nicola.paltrinieri@ntnu.no

Hydrogen has the potential to channel a large amount of renewable energy from the production sites to the end users. Nevertheless, safety aspects represent the major bottleneck for its widespread utilization. The knowledge of past hydrogen-related undesired events is fundamental to avoid the occurrence of similar accidents in the future. Databases such as HIAD 2.0 and H_2Tools are dedicated to those accidents, but the scarcity of structured and quantitative information makes it difficult to apply advanced data-driven analyses based on Machine Learning (ML). In this paper, undesired events related to the hydrogen value chain were selected from the HIAD 2.0 and MHIDAS databases. These records were collected in a structured repository tool, namely Hydrogen-related Incident Reports and Analyses (HIRA). The definition of its features is based on a critical comparison of the primary reporting systems, and an analysis of the literature regarding H_2 safety. Subsequently, text mining tools were used to analyze the event descriptions in natural language, extract relevant information and data, and sort them in the database. Finally, the new database was analyzed through Business Intelligence (BI) and ML classification tools. Data-driven analyses could help identifying valuable information about H_2-related undesired events, promoting a safety culture, and improving accident management in the emerging hydrogen industry.

Keywords: Hydrogen safety, Incident reporting system, Accident analysis, Learning from accident, Decarbonization, Risk prevention, Safety management.

1. Introduction

Hydrogen has been indicated by the European Commission (EC) as one of the most promising energy carriers to reduce greenhouse gas emissions and make the transport sector and power generation environmentally sustainable in the forthcoming years (EC, 2018). It can be produced by water electrolysis (green hydrogen) or by methane steam reforming coupled with carbon capture and storage (blue hydrogen) and used in fuel cell systems with elevated efficiency and near zero pollutant emissions (Ustolin et al., 2022). Despite its inherent advantages, H_2 has several safety issues related to its peculiar physio-chemical properties (i.e., the broad flammability range and the low ignition energy) and its ability to permeate and embrittle most metallic materials. These properties have caused several undesired events in the past, with severe consequences on humans, equipment, and environment. The knowledge of these accidents and a deep understanding of their root causes are fundamental to avoiding similar events in the future.

Hence, safety reporting systems are necessary to collect information on H_2-related incidents, accidents, and near-misses, thus maximizing the lessons learned and the return of experience from their analysis. Machine Learning techniques can facilitate this process thanks to their capability of dealing with large datasets, extracting relations among attributes, and predicting incident outcomes in terms of fatalities, injuries, and financial losses (Stefana et al., 2022). Dedicated databases for H_2-related events, such as the Hydrogen Incident and Accident Database (HIAD 2.0) and the Hydrogen Incident Reporting Tool (H_2Tools), are already publicly available. They provide meaningful information for classical statistical analyses, and, in some cases, they offer in-depth cause investigations. Nevertheless, the scarcity of quantitative information, the lack of a common taxonomy, and ambiguous features' definitions represent important limitations. These drawbacks make it difficult to apply advanced data-driven analyses based on ML.

In light of this, the present study aims at creating a systematic database for hydrogen-related accidents, which allows the adoption of ML tools to make predictions useful for enhancing hydrogen safety management. Firstly, the existing safety reporting systems, whether hydrogen-specific or generic, have been studied to analyze their strengths and limitations. Secondly, a new database, namely Hydrogen-related Incident Reports and Analyses, has been created by selecting proper features and defining the categories unambiguously. Hence, the database has been studied through Business Intelligence (BI) tools.

1.1. *Existing accident databases*

Several structured databases for major industrial accidents are already available. The French database Accident Reporting Information Analysis (ARIA) (BARPI, 2023) collects all types of events which are considered dangerous to human health, environment, or public safety, and contains 395 H_2-related events. The European Major Accident Reporting System (eMARS), created by the EC Joint Research Centre (JRC), includes 96 hydrogen releases and near misses (EC, 2023). The British Major Hazard Incident Data Service (MHIDAS) contains 104 hydrogen-related events which resulted in an ”offsite” impact. The accident database of the Institution of Chemical Engineers (IChemE, 2023) provides a brief description and the root cause analysis of 40 hydrogen releases, which occurred in the chemical and process industry. The National Response Centre (NRC) (EPA, 2023) is an American accident database that has been regularly updated from 1981 to 2001 and includes 120 H_2-related events. The Failure and Accidents Technical Information System (FACTS) collects detailed information on more than 481 hydrogen releases (Campari et al., 2023).

In addition, three H_2-specific safety databases are already in place and another one is under development. HIAD 2.0 is a multi-use platform to derive information for risk assessment and lessons learned. It includes a data entry form with a combination of narrative fields and predetermined options. The lack of quantitative information only allows for deriving high-level conclusions and general best practices (Wen et al., 2022). H_2Tools is curated by the Pacific Northwest National Laboratories (PNNL) and collects high-quality reports to get an understanding of the event, its root causes, consequences, and lessons learned. Data fields with predetermined options are often too generic to create reliability data useful for quantitative risk assessment. The National Renewable Energy Laboratory (NREL) collects failure and maintenance data from 44 hydrogen fueling stations and publishes anonymous reports with data regarding safety, maintenance, and reliability of industrial equipment. However, operating conditions and failure causes are partially missing and not adequately investigated. Finally, the Center for Hydrogen Safety (CHS) is developing a tool for reporting hydrogen equipment failures and obtaining failure rates specific to these components. The form includes information regarding the involved equipment, H_2 state, consequence, system response, and mitigation strategies. Nevertheless, operating conditions, location within the plant, component age, and failure mode are not considered. Hence, it is not possible to attribute a component failure rate to a specific failure mode (West et al., 2022).

Table 1 summarizes the main characteristics of the existing H_2 safety databases.

Table 1. Comparison of the existing hydrogen safety data collection tools

General feature	HIAD 2.0	H_2Tools	NREL CDPs	CHS Failure Rates
Records collected	655 events	221 events	44 fueling stations	Unknown
Year of creation	2004	2006	2020	Under development
Public access to data	Yes	Yes	Partial	Unknown
Initiating event	Yes	Yes	Yes	Yes
Location	Partial	No	Yes	Yes
Type of application	Yes	Yes	Yes	Yes
Primary cause	Yes	Yes	Yes	No
Contributing causes	Partial	Yes	Yes	No
Release size	Yes	No	Partial	Yes
Operating conditions	Yes	Yes	Yes	Partial
Components involved	Yes	Yes	Yes	Yes
Component size	Partial	No	Partial	Yes
Site inventory	No	No	Yes	Partial
Confinement	Yes	No	No	Yes
Failure symptoms	No	Yes	No	No
H_2 detection	No	No	Yes	Partial
Ignition	Yes	Yes	No	Yes
Consequences	Yes	Yes	Yes	Partial
Mitigation systems	No	No	No	Yes
Lessons learned	Yes	Yes	Yes	No
Recurrence frequency	No	Yes	No	No
Maintenance	No	No	Yes	Partial
Regular reporting	No	No	Yes	Yes
Data quality check	No	No	Yes	Unknown

2. Methodology

A new structured database of H_2-related undesired events has been created to achieve the objective of this study. HIAD 2.0 and MHIDAS were selected as primary sources to collect the relevant records. The former represents the main source for hydrogen accidents, while the latter has been chosen because its records are not included in HIAD 2.0. Since MHIDAS is a generic data source, we filtered the events by searching for the keyword “hydrogen” in the released substance field and in the event description. Among these records, there are incidents or accidents classified as H_2-related ones although hydrogen was not directly responsible for the undesired event. Consequently, we defined a set of inclusion and exclusion criteria to ensure the quality and consistency of the data. For instance, records that occurred in the H_2 supply chain (e.g. production, storage, transportation, and utilization) or involved hydrogen-specific technologies or other equipment dedicated to H_2 (e.g. flanges, piping, compressors, etc.), were deemed relevant along with releases of pure H_2 or H_2-rich mixtures. On the contrary, records concerning refineries (e.g., desulphurization units), or chemical plants for chlorine and ammonia production were excluded, as well as events in which H_2 was formed as a by-product of unwanted reactions, or car accidents not directly caused by the fuel-cell powertrain. The inclusion and exclusion criteria were applied by reading the records’ full descriptions and analyzing further available information. The included events were merged into a unique database. Dates, places, primary event IDs, sources, and full event descriptions were compared to identify and remove any duplicates. Then, the data were structured through a set of features, representing the columns of the database. A feature, whether numerical, categorical or textual, is a “measurable property of an object or event with respect to a set of characteristics”, and provides a machine-readable way to describe the relevant objects (ISO, 2022). For each feature, labels indicating the classes of the object under investigation were defined (Tamascelli et al., 2020).

Incident records usually have textual descriptions, which contain a large amount of hidden information. For this reason, we used Text mining (TM) to extract useful information from unstructured or semi-structured sources through the identification of trends and patterns (Feldman and Sanger, 2007). In this case, it was required to eliminate unwanted elements and redundancies to extract only the relevant information. The text preprocessing comprehended text reduction, transformation, filtering, and normalization. In addition, the text was decomposed into tokens (i.e., words or phrases), which represent the key elements for the classification of the features of the new database. The objective was to automate the database completion by mining information from the textual descriptions of the events. The methodology to create the HIRA database is depicted in Figure 1.

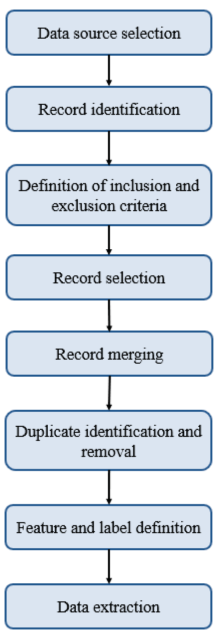


Fig. 1. Methodology for the incident database creation

The effectiveness of the TM process was tested by assigning the correct category of certain features to each event through machine learning tools. Three models were used for this classification task, optimized, and compared: Artificial Neural Network (ANN) (Štohl and Stibo, 2019), Logistic Regression (LR) (Maalouf, 2011), and Random Forest (RF) (Fawagreh et al., 2014). ANN was optimized with 70 neurons in hidden layers, ReLU activation function, and 60 maximum iterations; the Adam optimization function was used as a solver. C parameter equal to 550 and Ridge regularization were set for the LR model. Finally, RF was optimized with 48 trees and 65 subsets of features.

Then, Business Intelligence tools were used to analyze the structured information collected into the HIRA database. BI uses the Extract-Transform-Load (ETL) method to aggregate and systematize structured and unstructured data within a unique data management system (Microsoft, 2023). In this case, the key steps of the ETL process are (Trujillo and Luján-Mora, 2003):

- selecting the sources for extraction (i.e., HIAD 2.0 and MHIDAS databases);
- getting the data from the source location;
- validating and cleaning the data by introducing predefined inclusion and exclusion criteria;
- transforming the sources by systematizing the incident records through new features and labels;
- joining the sources of information in the HIRA database;
- selecting the target to load (e.g., the operational status);
- mapping source attributes to target attributes.

BI tools allowed creating dynamic dashboards, charts, graphs, and maps of data to visualize the results.

3. Results and discussion

By combining the 630 events collected in HIAD 2.0 and the 104 H_2-related accidents retrieved from MHIDAS, we obtained a preliminary set of 734 events. The application of inclusion and exclusion criteria, and the identification of duplicates allowed creating a new database with 325 events. The labels were defined to ensure clarity and avoid redundancies and repetitions. The inclusion of structured information into newly defined categorical fields allowed automatized analyses through BI and the adoption of artificial intelligence tools. For clarity, Table 3 summarizes the structure of HIRA.

Certainly, the compilation of the fields characterizing the database is time-consuming and requires a great deal of effort. For this reason, text-mining tools, aided by machine learning algorithms, were employed to automate the process of cataloguing information into the database. The tokens were used for the database compilation, which represents a classification problem. We trained LR, ANN, and RF models for this purpose. Then, three metrics, i.e., accuracy, precision, and recall, were used to evaluate the performance of these three classifiers. Accuracy indicates the fraction of correct predictions (Eq. 1), precision refers to the fraction of true positive predictions (Eq. 2), and recall expressed the fraction of positive labels correctly predicted (Eq. 3) (Seliya et al., 2009).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

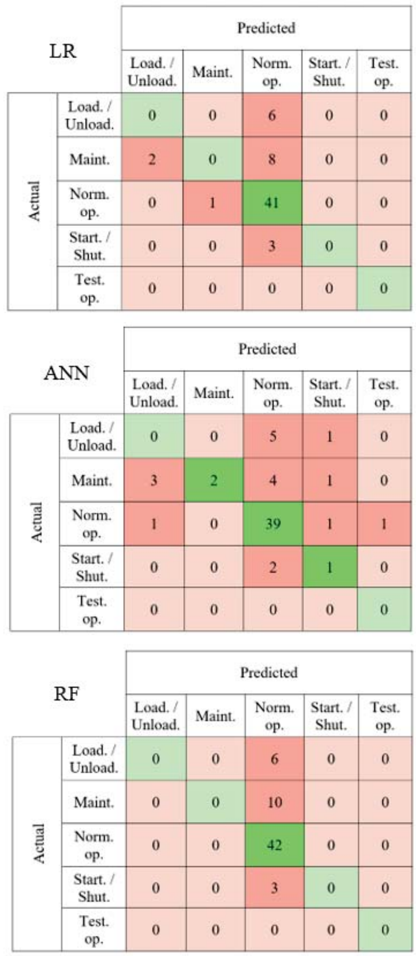
$$Recall = \frac{TP}{TP + FN} \tag{3}$$

where TP indicates true positives, TN the true negatives, FP the false positives, and FN the false negatives.

We classified various features but, in this study, we included only the operational status of the facility at the time of the event. This feature specifies the type of operational condition (normal, maintenance, startup, shutdown, etc.) of the facility when the event occurred. Accuracy, precision, and recall for the three machine learning algorithms are summarized in Table 2. In addition, the confusion matrices for the three model are shown in Figure 2.

Table 2. Evaluation metrics for the ML algorithms

	Accuracy	Precision	Recall
LR	0.672	0.487	0.672
ANN	0.689	0.713	0.689
RF	0.689	0.474	0.689



LR

Actual \ Predicted	Load. / Unload.	Maint.	Norm. op.	Start. / Shut.	Test. op.
Load. / Unload.	0	0	6	0	0
Maint.	2	0	8	0	0
Norm. op.	0	1	41	0	0
Start. / Shut.	0	0	3	0	0
Test. op.	0	0	0	0	0

ANN

Actual \ Predicted	Load. / Unload.	Maint.	Norm. op.	Start. / Shut.	Test. op.
Load. / Unload.	0	0	5	1	0
Maint.	3	2	4	1	0
Norm. op.	1	0	39	1	1
Start. / Shut.	0	0	2	1	0
Test. op.	0	0	0	0	0

RF

Actual \ Predicted	Load. / Unload.	Maint.	Norm. op.	Start. / Shut.	Test. op.
Load. / Unload.	0	0	6	0	0
Maint.	0	0	10	0	0
Norm. op.	0	0	42	0	0
Start. / Shut.	0	0	3	0	0
Test. op.	0	0	0	0	0

Fig. 2. Confusion matrices for the operational status classified through LR, ANN, and RF models

ANN outperforms the other algorithms in terms of precision, while both Logistic Regression and Random Forest have very low performances. In addition, the confusion matrices highlight how LR and RF are unable to correctly predict any minority class (i.e., operational status different from ”Normal operations”). On the other hand, ANN cannot label correctly the ”Loading/unloading operations” and ”Testing operations”, but it is more accurate in predicting ”Maintenance” and ”Startup/shutdown”. Nevertheless, the overall performance of this model cannot be considered fully satisfactory. Machine learning algorithms can make reliable predictions when trained on a large number of input data. Unfortunately, the limited operational experience with hydrogen along with the low market penetration of hydrogen-specific components results in a low number of H_2-related incidents. Another factor that further reduces the amount of safety data is associated with the regulatory framework regarding incident and accident reporting. In the EU countries and the United States, there are binding requirements to report to the competent authorities all industrial accidents involving dangerous substances to develop cause analyses and safety investigations. But this is not the case for most Asian, African, and South American countries, thus leading to a general under-reporting of H_2-related undesired events (Campari et al., 2023). Despite these limitations, better performances will be achieved by including more sources of information in the HIRA database, such as H_2Tools or the NREL CDPs. Furthermore, we can expect that more complete and detailed event descriptions will be reported in the future, thus providing a larger set of examples for the combined application of TM and ML tools. Finally, the HIRA database could be complemented with additional features related to the existing safety culture of the organization involved in the event. For example, a new feature for the previous accidents at the facility can be included, considering the information provided by the NREL database.

The analysis of the database through BI tools has proven that most H_2-related events (i.e., 202 out of 325) occurred during the normal operations of the plants. This is a predictable outcome since an industrial facility operates under ”normal conditions” for most of its working life. It is interesting to note how 14.5% of the total events occurred during maintenance activities, despite they are normally carried out every five years (in the case of hydrogen refueling stations and related equipment) (ISO, 2020) or in extraordinary circumstances. These findings are shown in Figure 3.

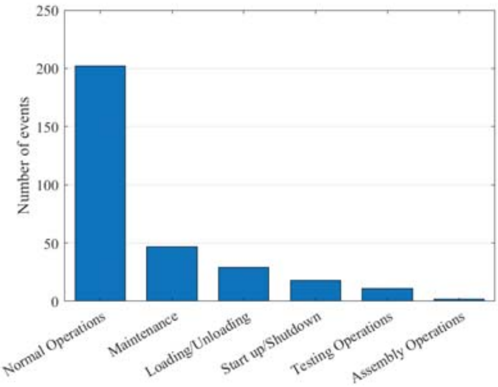


Fig. 3. Operational status of the facility at the time of the event

Most industrial components for hydrogen handling and storage have not been designed specifically for H_2 applications or have been used for other substances and adapted for hydrogen. Predictive maintenance of such components is of the utmost importance to minimize the risk of equipment failure and avoid the loss of containment. Despite this, the existing standards and guidelines for inspection and maintenance planning are poorly adapted to equipment operating in H_2 environments. For instance, the risk-based inspection and maintenance methodology does not consider most hydrogen-induced material failures, thus increasing the uncertainty and affecting the decision-making process (Campari et al., 2022). In addition, the recommended maintenance actions are often indicated by the equipment's producers, in the absence of a unified regulatory framework. Under these conditions, the success of maintenance operations mostly depends on the experience and training level of the operators.

4. Conclusions

This study proposes an approach based on TM, ML, and BI tools for learning from H_2 accidents.

Such approach is grounded on the creation of a new database, called HIRA, containing 325 undesired events that involved hydrogen technologies or, more generally, occurred in the H_2 value chain. HIRA is characterized by 43 numerical, categorical, or textual features and related labels, which refer to different characteristics of items and components, pieces of information about the accident scenario, and quantitative data about physio-chemical properties of H_2. The database was populated by applying TM tools, while three machine learning techniques (i.e. LR, ANN, RF) were employed to classify and predict relevant accident features. For space constraints, this paper only reports and discusses the results about operational status of the facility at the time of the event. Finally, BI tools permitted creating dynamic dashboards to visualize the results. This approach facilitates critical analyses of H_2-related events, thus allowing lessons learned for safely managing H_2 and the related components, technologies, and plants. The consideration of other data sources and the introduction of other accidents could further enhance our approach and its results.

Acknowledgement

This work was undertaken as a part of the research project SUSHy ("Sustainability and Cost-Reduction of Hydrogen Stations through Risk-Based, Multidisciplinary Approaches"), and the authors would like to acknowledge the financial support of the European Interest Group (EIG) CONCERT-Japan (Grant No. 334340).

References

BARPI (2023). Accident Reporting Information Analysis.

Campari, A., M. A. Darabi, F. Ustolin, A. Alvaro, and N. Paltrinieri (2022). Applicability of risk-based inspection methodology to hydrogen technologies: A preliminary review of the existing standards. *Proceedings of the 32nd European Safety and Reliability Conference*.

Campari, A., A. Nakhal, F. Ustolin, A. Alvaro, A. Ledda, P. Agnello, P. Moretto, R. Patriarca, and N. Paltrinieri (2023). Lessons learned from HIAD 2.0: Inspection and maintenance to avoid hydrogen-induced material failures. *Comput. Chem. Eng. 173*.

EC (2018). *In-depth analysis in support of the Commission Communication COM(2018) 773: A Clean Planet for all - A European strategic long-term vision for a prosperous, modern, competitive and climate neutral economy*.

EC (2023). European Major Accident Reporting System Database.

EPA (2023). National Response Center Database.

Fawagreh, K., M. Gaber, and E. Elyan (2014). Random forests: From early developments to recent advancements. *Syst. Sci. Control Eng. 2*.

Feldman, R. and J. Sanger (2007). *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*. Cambridge University Press.

IChemE (2023). Institution of Chemical Engineers Database.

ISO (2020). *ISO 19880-1 - Gaseous hydrogen fueling stations - General requirements*.

ISO (2022). *ISO/IED 23053:2022 - Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)*.

Maalouf, M. (2011). Logistic regression in data analysis: an overview. *Int. J. Data Anal. Tech. Strateg. 3*.

Microsoft (2023). What is business intelligence?

Seliya, N., T. Khoshgoftaar, and J. V. Hulse (2009). A study on the relationships of classifier performance metrics. *Proceedings of the 21st IEEE International Conference on Tools with Artificial Intelligence*.

Stefana, E., D. Ghidoni, F. Fanizza, F. Marciano, P. Cocca, and N. Paltrinieri (2022). Towards machine learning application for safety in confined spaces: Creating an incident database. *Proceedings of the 32nd European Safety and Reliability Conference*.

Tamascelli, N., N. Paltrinieri, and V. Cozzani (2020). Predicting chattering alarms: A machine learning approach. *Comput. Chem. Eng. 143*.

Trujillo, J. and S. Luján-Mora (2003). A UML based approach for modeling ETL processes in data warehouses. *Proceedings of the 22nd International Conference on Conceptual Modeling*.

Ustolin, F., A. Campari, and R. Taccani (2022). An extensive review of liquid hydrogen in transportation with focus on the maritime sector. *J. Mar. Sci. Eng. 10*.

Wen, J., M. Marono, P. Moretto, E. Reinecke, P. Sathiah, E. Studer, E. Vyazmina, and D. Melideo (2022). Statistics, lessons learned and recommendations from analysis of HIAD 2.0 database. *Int. J. Hydrog. Energy 47*.

West, M., A. Al-Douri, K. Hartmann, W. Buttner, and K. Groth (2022). Critical review and analysis of hydrogen safety data collection tools. *Int. J. Hydrog. Energy 47*.

Štohl, R. and K. Stibo (2019). Predicting safety solutions via an artificial neural network. *IFAC - PapersOnLine 52*.

Table 3. Features and labels in the HIRA database

Features	Type	Labels
Event ID	Categorical	ID 2 (ID 11 in HIAD); ID 5 (ID 2467 in MHIDAS).
Location	Categorical	America, United States, Richmond; Europe, Germany, Berlin.
Date	Numerical	10/04/1989; 10/03/1978.
Full description	Textual	Initial hydrogen leak resulted from the failure of an elbow welded on the pipeline body [...] the reactor collapsed damaging nearby equipment.
Application stage	Categorical	H_2 Production; H_2 Transportation; H_2 Storage; H_2 Utilization.
Detailed application	Categorical	Aerospace; Automotive; Chemical and petrochemical; Power generation; Laboratory; Maritime; etc.
Location	Categorical	Airport; Compression room; Laboratory; Loading area; Refueling station; Storage area; etc.
Primary cause	Categorical	Component failure; Design error; External cause; Installation error; Maintenance error; Operational error.
Secondary cause	Categorical	Absence of safety systems; Excessive temperature; H_2 accumulation; H_2 embrittlement; Overpressure; Presence of ignitable mixtures; etc.
Primary event	Categorical	H_2 condensation; H_2 leakage; Formation of explosive mixture; etc.
Ignition	Categorical	Yes; No.
Final scenario	Categorical	Dispersion; Explosion; Flash fire; Jet fire; Near miss; Safe release; etc.
Secondary consequences	Categorical	Ejected debris; H_2 tank destruction; Pipeline rupture; Secondary fires; etc.
Domino effect	Categorical	Yes; No.
Primary item	Categorical	Airship; Battery; Boiler; Check valve; Compressor; Cylinder; Electrolyzer; Flange; Heat exchanger; Pipeline; Safety valve; Tank; etc.
Prim. item dimensions	Numerical	Volume: 42 m^3, Diameter: 610 mm, Width: 6.3 mm; etc.
Prim. item material	Categorical	Austenitic steel; Brass; Carbon steel; Copper; C-0.5Mo steel; etc.
Prim. item pressure	Numerical	25 bar; 75 bar; 200 bar; 241 bar; 500 bar; etc.
Secondary item	Categorical	Same as the primary item.
Sec. item dimensions	Numerical	Same as the primary item dimensions.
Sec. item material	Categorical	Same as the primary item material.
Sec. item pressure	Numerical	Same as the primary item pressure.
Substance released	Categorical	H_2; H_2/NH_3 mixture; H_2/N_2 mixture; H_2/O_2 mixture; Syngas; etc.
Other substances	Categorical	Same as the substance released.
Storage medium	Categorical	Gas; Liquid.
Storage quantity	Numerical	0.5 m^3; 6 m^3; 70.4 m^3; 510 m^3; 20,000 m^3; etc.
Location type	Categorical	Open space; Semi-confined; Confined.
Actual pressure	Numerical	Same as primary item pressure.
Actual temperature	Numerical	443 K; 533 K; 653 K; 714 K; 770 K; etc.
Operational status	Categorical	Assembly operation; Inspection; Loading; Maintenance; Normal operation; Shutdown; Start-up; Testing operation; Unloading.
Release type	Categorical	Gas; Liquid; Mixed.
Release quantity	Numerical	Same as the storage quantity.
Active barriers	Categorical	Alarm; H_2 detector; Flare; Safety valve; Sensor; Sprinkler system; etc.
Passive barriers	Categorical	Containment basin; Reactor steel skirt; Rupture disk.
Procedural barriers	Categorical	Firefighters intervention; Emergency plan; Emergency shutdown; Evacuation; Feed blocked; Fire extinction; Road closed; Ventilation; etc.
Fatalities	Numerical	1; 3; 6; 19; 30; 62; etc.
Injured persons	Numerical	Same as the fatalities.
Evacuees	Numerical	Same as the fatalities.
Environmental damage	Categorical	Yes; No.
Economic loss	Numerical	0.1 M USD; 0.1 M USD; 0.1 M USD; 0.1 M USD; 0.1 M USD; etc.
Hole diameter	Numerical	1 mm; 2 mm; 10 mm; 25 mm; 48 mm; etc.
Hole shape	Categorical	Circular; Crack; Gap.
Post-event notes	Textual	The company involved has implemented revised procedures [...] to ensure that it can only be inserted in the correct orientation.