HOLY: An Ontology covering the Hydrogen Market

Kiara M. Ascencion Arevalo[1[0000-0002-4158-9478]], Christoph Neunsinger[[0000-0001-9312-208X]], Roland Zimmermann[1[0000-0001-7380-5908]], Ralph Blum[1], Kendra Weakly[1[0009-0002-4224-4153]],

[1] Technische Hochschule Georg Simon Ohm, Bahnhofstraße 87, 90402 Nürnberg, Germany
kiaramarnitt.ascencionarevalo@th-nuernberg.de

Abstract. This paper presents the *Hydrogen Ontology (HOLY)*, a domain ontology modeling the complex and dynamic structures of hydrogen-based markets. The hydrogen economy has become a politically and economically crucial sector for the transition to renewable energy, accelerating technological and socio-economic innovations. However, the attainment of market insights requires a large variety of informational concepts which are predominantly found in unstructured text data. *HOLY* provides the necessary structure for the representation of these concepts. Through a top-down approach, *HOLY* defines taxonomies based on a hierarchical structure of products and applications. In addition, to ensure reusability, the ontology incorporates components from established ontologies in its structure. As a result, *HOLY* consists of over 100 classes defining information about organizations, projects, components, products, applications, markets, and indicators. Hence, our work contributes to the systemic modeling of the hydrogen domain with a focus on its value chain. Formally, we represent and validate the ontology with Semantic Web Technologies. *HOLY* includes lexical-semantic information (e.g., synonyms, hyponyms, definitions, and examples) to simplify data integration into knowledge acquisition systems. Therefore, we provide a foundation for the retrieval, storage, and delivery of market insights. A first application based on *HOLY* at the Fraunhofer IIS offers an up-to-date market overview of developments in the fuel cell environment.

Keywords: Hydrogen Ontology, Market Modeling, Ontology Engineering, Ontology-based Information Extraction, Value Chain Representation, PEM Fuel Cell.

1 Introduction

The hydrogen market is currently undergoing rapid changes. It is predicted to nearly double in size from 170 billion USD in 2021 to about 300 billion USD by 2027 [1], and - being a possible source for the green energy transition - it obtains strategic importance for nations. The most significant contributor is the European Union (EU) with its EU Hydrogen Strategy[1]. However, other countries such as China, India, and the United States are also working to push this sector forward [2]. Consequently, the hydrogen sector is becoming attractive to companies investing in hydrogen projects. It also raises

[1] cf., https://energy.ec.europa.eu/topics/energy-systems-integration/hydrogen_en

challenges concerning competition, development, and identification of key technologies to overcome current infrastructure limitations and political obstacles [3]. Hence, industrial, governmental, and research institutions require continuous monitoring of market developments to assess and tune their strategies.

In order to monitor developments in such a dynamic market, several aspects have to be addressed. Firstly, dynamically-changing market structures (e.g., the emergence of new stakeholders or new relations between existing stakeholders) as well as technological developments (e.g., improvements of specific fuel cell [FC] types) must be modeled in relation to each other. Secondly, these types of insights come from varying sources (e.g., publications [4], magazines [2], public institutions[2], company websites[3], and newspapers[4]). In other words, it is required to structure the necessary information (i.e., which data entities and data relationships are of interest) and enable a continuous process of retrieval (R), storage (S), and delivery (D) for the above-mentioned institutions.

To represent heterogeneous information (e.g., about markets and technology), domain ontologies capturing the conceptual patterns of a domain have been proposed in literature (e.g., in business [5], for infrastructure [6], or in the technology domain [7]). Ontologies provide a common representation of the structure of information in a particular domain by defining a shared vocabulary to enable people and software agents to share information [8].

Existing proposals are, to the authors' knowledge, sparse and insufficient in capturing the relevant information in the hydrogen domain. Therefore, the research question, *'How can knowledge about hydrogen-related microeconomic systems (domains) be modeled to enable extraction of market insights?'* guided our research.

We propose the domain-specific *Hydrogen Ontology (HOLY)* as a structural backbone for the hydrogen sector's R, S, and D processes, serving as a continuously-growing knowledge base for strategic foresight purposes. The development of *HOLY* is based on the *Linked Open Terms (LOT)* approach [9], an established methodology in the semantic literature used in different domains (e.g., in agriculture [10], information and communication technology [11], environmental management and sustainability [12], and in industrial context [13]). *HOLY* is already being used by the Fraunhofer Institute for Integrated Systems (IIS) for R, S, and D of market insights in the Atlant-H[5] project. The project aims to create a tool for automatically analyzing international market activities in the hydrogen environment. Hence, it employs natural language processing (NLP) for retrieval R and a graph database for storage S. Delivery D is based on SPARQL, a query language for the Resource Description Framework (RDF). Query results are presented in a user-friendly business intelligence front-end[6]. The ontology advances the understanding of information conceptualization in a large, dynamically-changing market while considering an automated process (R, S, D). At the same time,

---

[2] cf., https://energy.ec.europa.eu/topics/energy-systems-integration/hydrogen_en

[3] cf., https://www.ballard.com/fuel-cell-solutions/fuel-cell-power-products

[4] cf., https://www.theguardian.com/environment/2018/sep/17/germany-launches-worlds-first-hydrogen-powered-train

[5] Atlant-H: https://www.scs.fraunhofer.de/de/referenzen/atlant-H.html

[6] Atlant-H Front-end Demo: https://tinyurl.com/yr964ycu

HOLY provides a structured, yet dynamically-adaptable market model for the business community in the hydrogen sector.

Chapter 2 introduces related work, focusing on the usage of ontologies to model technical and market-related subjects and as enablers for the extraction of knowledge. Our approach for *HOLY* is detailed in chapter 3, where we follow *LOT*'s set of structured development steps. We conclude with a summary and an outlook on future work in chapter 4.

2 Related Work

Using domain ontologies to satisfy the analytical needs of market participants is a widespread practice among institutions. The EU used ontologies in the early 2000s within the Environmental Data Exchange Network for Inland Water (EDEN-IW) project to model information on water quality by integrating heterogeneous databases from governments [14]. The Market Blended Insight (MBI) project offered an approach to model a Business-to-Business (B2B) market in a closed domain setting with seven key partners. It includes market models in its ontology engineering process [5]. Domain ontologies position themselves as a recurring solution for areas requiring conceptual modeling and integrating information from different data sources [15].

In the hydrogen domain, knowledge graphs have been used to facilitate knowledge representation and discovery in the field of scientific research [16–18]. Additionally, interest in monitoring information pertaining to the hydrogen market can be seen in the creation of databases for hydrogen projects[7,8], and hydrogen companies[9]. To monitor the hydrogen market, the European Union realizes a rather broad approach based on the Tools for Innovation Monitoring (TIM) toolset [19, 20]. The interactive tool uses graph-based visualizations; however, these are not based on knowledge graphs or ontologies, but on the co-occurrence of keywords in the same texts [19, 20]. Nevertheless, interesting insights (e.g., into weak signals) are presented in curated reports (for 2021, see [21]).

For ontology development, reusing and building on existing, established ontologies is essential. *The Organization Ontology* was created in 2010 as a core ontology for company information, providing a logical structure for linked company data across numerous domains which includes roles and memberships [22]. The *Registered Organization Vocabulary* expanded upon *The Organization Ontology* in 2013 by extending the formal organization section with unique properties for displaying legal status and economic activities [23]. These activities can be classified using internationally accepted standards such as the Nomenclature of Economic Activities (NACE) [24], the Standard Industrial Classification [25], or the International Standard Industrial Classification [26]. In 2020, the *euBusinessGraph* added depth on registered businesses in the European Union by focusing on information concerning the harmonization of a company's legal type, its registered economic activity, geographic location, founding date,

---

7 cf., https://www.iea.org/data-and-statistics/data-product/hydrogen-projects-database
8 cf., https://commodityinside.com/reports/global-green-hydrogen-projects-database/
9 cf., https://www.enerdata.net/research/h2-database.html

classification as a start-up, and classifications for being state-owned or independent using concepts from the *Registered Organization Vocabulary* [27].

Ontologies are also utilized for Ontology Based Information Extraction (OBIE). OBIE is the conceptual structuring of information coming from Information Extraction (IE) technologies [28]. Hence, OBIE systems utilize ontological structures to filter for relevant terms or enrich them by setting extracted information into context with other information. For example, Raza Rizvi et al. [29] use ontologies to enable generalization of their OBIE system and extraction of tabular information from documents irrespective of the domain. Furthermore, the MUSING IE architecture relies on an ontology to consolidate information from its different sources (e.g., company profiles from Yahoo! Finance, company websites, newspaper articles, and company reports) concerning knowledge about companies, region-specific insights, and economic indicators [28].

3 The Hydrogen Ontology (HOLY)

3.1 Methodology

This paper's methodology for ontology development is based on the *LOT* framework [9], as shown in **Fig. 1**. In the requirements specification stage, the ontology's requirements (i.e., the goal, scope, use cases, and functional requirements) are identified. The implementation stage is an iterative process where a conceptual model describing the problem and its solution is created; this model is formalized in an ontological language, such as RDF, to generate machine-readable models and including metadata. Additionally, existing ontologies are identified and integrated into the developed structure. In the publication stage, the ontology is made available online via its URI, in both human-readable documentation and machine-readable file formats. The maintenance stage aims to keep the ontology updated throughout its lifecycle. Hence, activities can be triggered during the development process or later on. As a supporting activity, knowledge acquisition is constantly present in the ontology development process. Knowledge acquisition[10] relies on a range of sources including interviews, publicly-available resources from websites, magazines, books, non-ontological and ontological types as well as best practices.

[10] HOLY Knowledge Acquisition Process: https://purl.org/holy/knowledge_acquisition

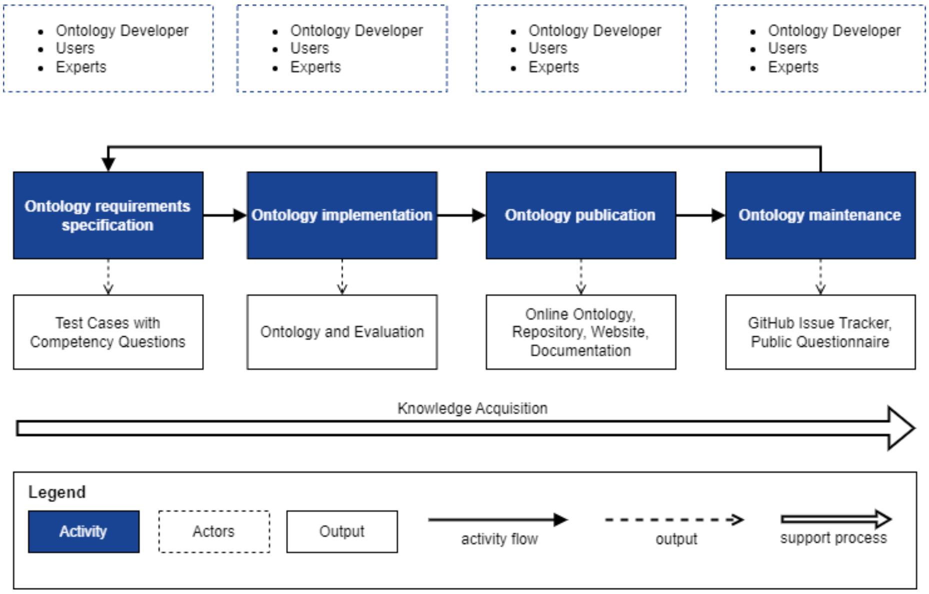

Fig. 1. The *Linked Open Terms (LOT)* Framework used to develop the *Hydrogen Ontology*.

The results of each stage are described in the following chapters. In addition to the *LOT* framework, guidance from *Ontology Development 101* by Noy & McGuinness [30] is applied to develop class definitions, hierarchies, properties, and class instantiations. Methods for knowledge acquisition follow the NeOn Methodology [31].

3.2 Ontology Requirements Specification

HOLY was developed within the Atlant-H hydrogen project to structure information from the hydrogen economy and serve as the backbone of an OBIE and NLP-powered text processing system for automatic market activity analysis. Hereby, *HOLY*'s objective is the representation of domain knowledge of the hydrogen sector to track developments in the market. Consequently, *HOLY* is intended to be used for decision-making, market monitoring, and to facilitate research planning for industry players, governmental institutions, and research institutions. In addition, *HOLY*'s lexical-semantic information and structure facilitate performing Natural Language tasks such as Named Entity Recognition and Question Answering, which are employed to extract information from natural language sources automatically. Aside from the requirements of the Atlant-H project, other potential use cases[11] have been identified. These applications extend to various additional stakeholders, such as technology providers, investors, product users, and educational institutions.

The specification of ontology requirements was carried out by utilizing different inputs and models gathered from the knowledge acquisition process. In order to cover

[11] HOLY Use Case Specifiation: https://purl.org/holy/use_case_specification

domain knowledge of the hydrogen sector, the ontology incorporates both market structures and technological knowledge of the hydrogen domain. The former requires concepts and relationships of market actors, their roles, and interactions, while the latter requires concepts and relationships of hydrogen technologies and their components.

Michael E. Porter's 'Five Forces' framework structure has been identified as an appropriate source for market structure conceptualization. Porter's 'Five Forces' is an established, domain-independent framework used to evaluate the competitiveness of industries. The framework is based on five forces in a market [32]. These forces are used to identify required structures and corresponding knowledge (e.g., economic activities coming from NACE and definitions coming from the Cambridge Dictionary).

Market insights are derived from the report 'Geopolitics of the Energy Transformation: The Hydrogen Factor' of the International Renewable Energy Agency (IRENA) [2], the Hydrogen Council [3], the European Commission[12], news articles, and company websites[13]. The Hydrogen Council is an initiative comprising 150 organizations that represent the global hydrogen value chain. Their Hydrogen Insight Report from 2022 reflects on the maturity level of the hydrogen market and defines key factors for sustainable industry-wide growth [3]. As such, the necessity of establishing a global coverage perspective with details on geolocations was revealed since major hydrogen players are distributed worldwide [2]. Furthermore, the importance of tracking projects and indicators such as investments or funding was identified [3].

The acquisition of technological knowledge included identifying and categorizing the hydrogen market and its underlying products, components, and composition. Information was derived from the IRENA report [2], leading to the requirement of modeling the structure as displayed in **Fig. 2**. The structure follows the energetical path of hydrogen by showing the value chain from production to the end-user.

[12] cf., https://energy.ec.europa.eu/topics/energy-systems-integration/hydrogen_en
[13] cf., https://www.ballard.com/fuel-cell-solutions/fuel-cell-power-products

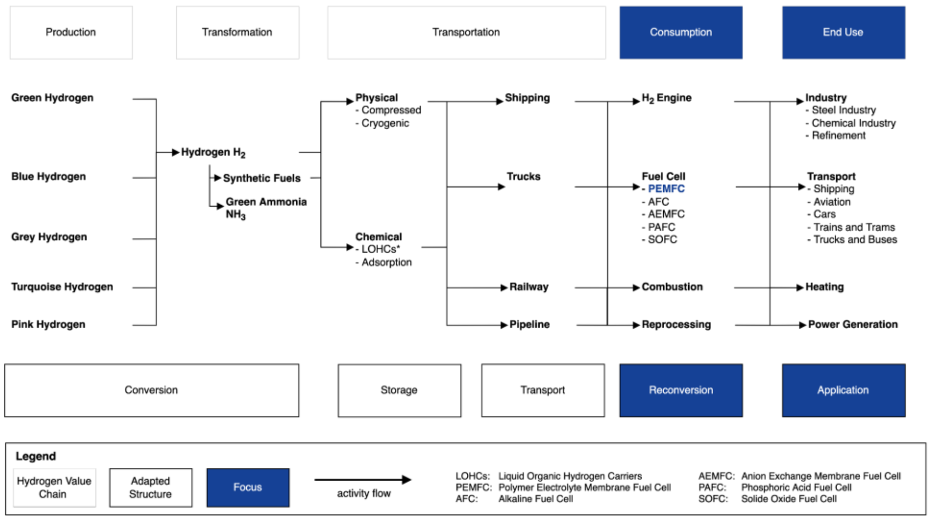

Fig. 2. Hydrogen Market Value Chain adapted and simplified from the IRENA report [2] and Information from TÜV Süd[14].

Concepts and definitions were extracted from the analysis of descriptions of governmental institutions[15], websites from active market participants[16], and scientific journals [32]. In order to properly represent said concepts for NLP tasks, it is required to include lexical-semantic data in the ontology. Therefore, synonyms, definitions, and examples for each class are included.

Every value chain stage of the hydrogen market can be detailed to a high granularity level. To exemplify detailed modeling of the different stages through *HOLY*, we extensively modeled the reconversion stage (i.e., reconversion of hydrogen into electrical power). More specifically, we focused on the established Proton-exchange Membrane Fuel Cell (PEMFC) technology, a sub-technology of fuel cells that is forecasted to receive the fastest commercial growth among fuel cell technologies in the next few years. At the same time, it is prognosed to stay in competition with other maturing and potentially upcoming fuel cell technologies [2].

We used test cases to formalize the functional ontology requirements identified. Test cases are formal descriptions of input, action, and outcomes used to ensure that a Linked Data resource conforms to a given set of requirements [33] - in this case, our ontology requirements specification document (ORSD)[17]. As such, they enable the evaluation of the quality and interoperability of *HOLY* by verifying that the terms are clearly and unambiguously defined [34]. Test cases were collected from industry participants as a set of competency questions as part of the Atlant-H project. We then defined our test cases based on the competency question in collaboration with Fraunhofer IIS internal

[14] cf., https://www.tuvsud.com/en/themes/hydrogen/explore-the-hydrogen-value-chain
[15] cf., https://www.energy.gov/eere/fuelcells/hydrogen-storage
[16] cf., https://www.ballard.com/fuel-cell-solutions/fuel-cell-power-products
[17] HOLY Requirements Specification: https://purl.org/holy/requirements

market experts and H2Ohm[18] hydrogen experts. Ontology requirements and corresponding test cases as either 'Technological' or 'Market' depending on the type of information to which they are related. A comprehensive list of our test cases can be found in our repository[19].

3.3 Ontology Implementation

In order to address the ORSD, two orthogonal dimensions - market and hydrogen technology - have to be conceptualized in the ontology.

The **market dimension** requires terms which enable the development of reusable abstract microeconomic structures. As a result, *HOLY* consists of the six main classes listed below and conceptualizes market relationships such as production, cooperation, geographic placement, and provision of goods or services.

- **Product:** delineates relevant and substitute technologies, products, and their components along the hydrogen value chain to cover the need for segmentation outlined by the market.
- **Application:** covers use cases for products and technologies under the categories of stationary and mobile applications.
- **Organization:** contains structural information about market participants and organization types following NACE.
- **GeographicMarket:** focuses on geographic units at the level of countries and continents. Smaller units (e.g., states or provinces) are mapped to these two classes.
- **Project:** provides a structure to classify project types by purpose and state and display their market role using object properties connected to the other five classes.
- **Indicator:** has connections to all other classes, allowing it to store performance information.

[18] Institut für Angewandte Wasserstoffforschung, Elektro- und Thermochemische Energiesysteme (H2Ohm):
https://www.th-nuernberg.de/einrichtungen-gesamt/in-institute/institut-fuer-angewandte-wasserstoffforschung-elektro-und-thermochemische-energiesysteme/

[19] HOLY Test Cases: https://purl.org/holy/test_cases

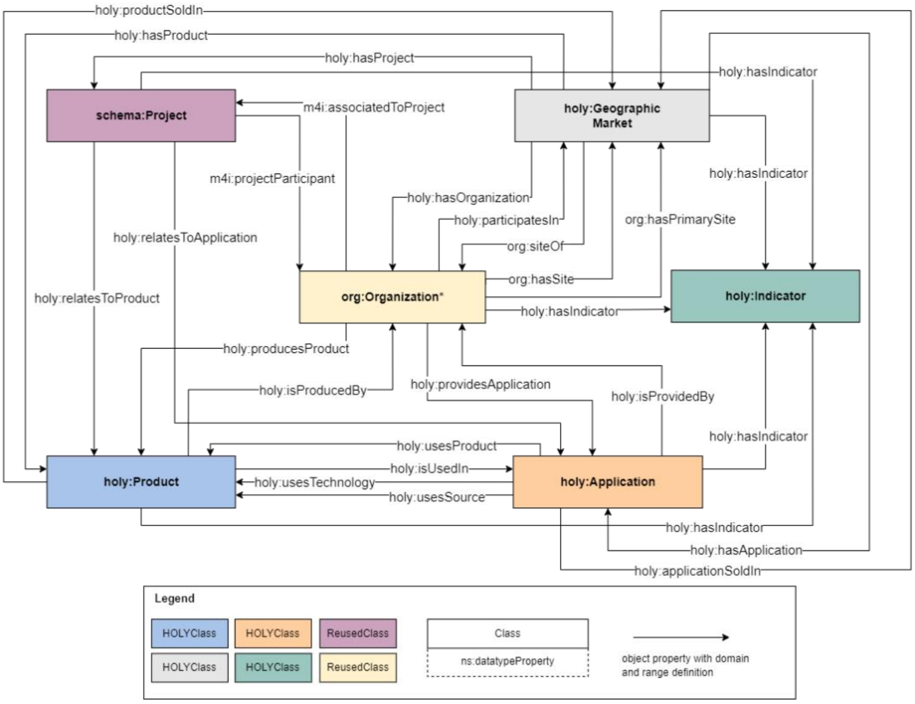

Fig. 3. Overview of the *Hydrogen Ontology*'s six main classes and related properties.

Fig. 3 provides an overview of the ontology schema including main classes and relationships. Organization and Project classes are defined by existing ontologies - *The Organization Ontology* (*org*) and *Schema.org* (*schema*) respectively - while the other classes belong to *HOLY*'s vocabulary. The subclasses of *org:Organization* and *holy:GeographicMarket* structure information to identify rivalry within the market based on segmentation by a player's economic activity and geographic location. The alignment with terms from official economic nomenclatures under *org:Organization* enables identification of buyers and suppliers. A threat of substitutes is addressed using the classes *holy:Product* and *holy:Application* by providing a structure to cluster PEMFCs and their substitutes according to their related applications. The class *schema:Project* provides a set of terms to model market players' efforts to develop the industry and classify them by objective (e.g., research-oriented, product-oriented, infrastructure-oriented or oriented to projects concerning circular economy) and state (finished, ongoing, or planned). The class *holy:Indicator* categorizes strategic information influencing the strength of a market participant and barriers faced by parties interested in competing in the hydrogen industry. Indicators are, for example, market share, investment, market size, and patents. Hence, the model covers the 'Market' requirements from the ORSD.

The **hydrogen dimension** specifies products and component classes related to the hydrogen domain and splits the market into segments along the stages of the hydrogen value chain (conversion, storage, transport, and reconversion). The separation between products and their components enables detailed insights into product composition. The modeling example of PEMFC technology embeds a more detailed level of granularity into the product taxonomy. A part of the hydrogen schema is shown in the following **Fig. 4**, illustrating different types of products on multiple levels of abstraction.

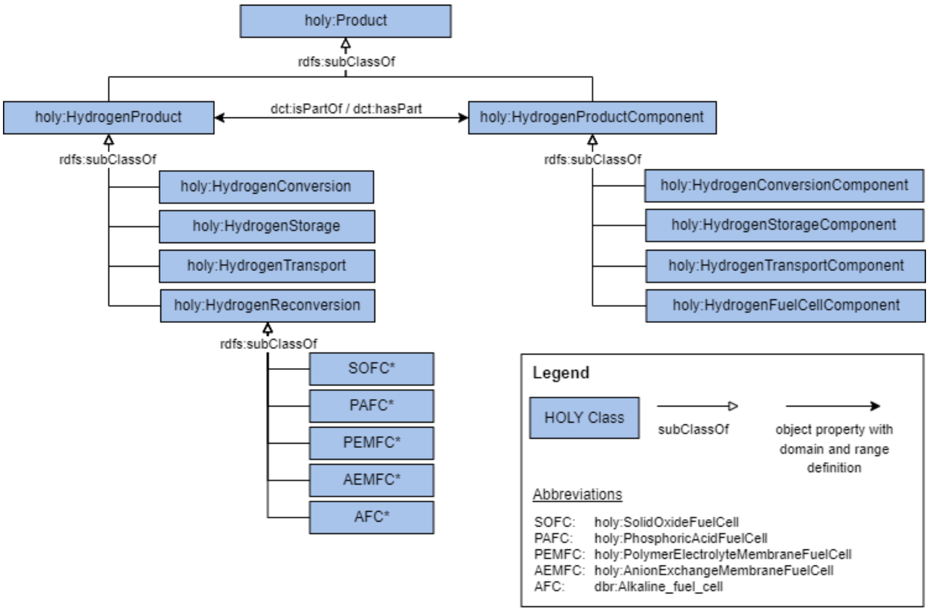

Fig. 4. A fragment from *HOLY*'s taxonomy focused on the *holy:HydrogenProduct* and *holy:HydrogenProductComponent* classes. A complete representation is available in the *HOLY* documentation[20].

With the purpose of ensuring the reusability and extensibility of the ontology, existing ontological concepts are integrated into *HOLY*'s structure. Following best practices, *HOLY* applies a multi-tiered approach to enable semantic interoperability using a range of ontologies - from top-level to low-level abstractions - across different domain sections [35]. The frameworks of *RDF Schema* (*RDFS*) and *Simple Knowledge Organisation System* (*SKOS*) act as top-level ontologies in *HOLY* and provide extended structures of hierarchical knowledge representation, establishing a common vocabulary. *RDFS* is used to define concepts and relationships while *SKOS* enables lexical-semantic information integration.

Mid-level ontologies cover areas of representation such as organizations, locations, and projects. *The Organization Ontology* provides a widely used and accepted framework for classifying organizations. *Schema.org*'s vocabulary delivers extensions to classify a project while providing structures for geographic segmentation. Lower-level ontologies extend these mid-level ontologies by building upon their class structures

[20] HOLY Documentation: https://purl.org/holy/doc

[35]. The *Registered Organization Vocabulary* (*rov*) extends *The Organization Ontology* regarding organizations' classifications and economic activities. At the same time, the *euBusinessGraph* (*ebg*) expands the *Registered Organization Vocabulary* even further towards additional company classification criteria such as defining whether an organization is state-owned, publicly traded, or a start-up [27]. *Schema.org* is enriched with the *Metadata4Ing* (*m4i*) ontology, which provides an extended information structure for identifying project participants and the status of projects. The *DBpedia Ontology* (*dbo*) offers geographic mapping through entity linking against its knowledge graph [36].

To facilitate OBIE, which utilizes NLP methods to extract information, natural language descriptors for the classes on both dimensions are necessary. To this end, an additional lexical-semantic layer is included in the ontology structure, explicitly declaring synonyms, abbreviations, keywords, examples, definitions, and definition sources. The main goal of this information is to provide a mapping between the words in source texts and the concepts represented in the ontology, allowing NLP techniques to identify the concepts being referred to. Hence, computational representations of the meaning of words in context are created including lexical relations between words in a language (e.g., synonyms or hyponyms). *HOLY* primarily uses the *SKOS* ontology and the *Wordnet RDF Ontology* (wn) to model information required for NLP pipeline processes. For example, *skos:altLabel* is used to capture class synonyms and abbreviations, *wn:hyponym* is used to represent relevant keywords that denote subtypes of a more general term, and *skos:hiddenLabel* is used for class-relevant keywords that can be used for text-based indexing in a more general context. In addition, *skos:example* is used to annotate specific examples such as companies (e.g., Plug Power Inc.), product names (e.g., FCwave™), or applications (e.g., Toyota Mirai) available on the market. The property *skos:definition* describes the terms used in the ontology, and the sources of these descriptions are given using the *dct:references* property.

The *Hydrogen Ontology* is built using OWL 2. *HOLY* combines newly created classes and properties with existing ontologies and vocabularies such as *The Organization Ontology*, *Schema.org*, and the *euBusinessGraph*. *HOLY* consists of 109 classes (6 main classes and 103 sub-classes), 35 object properties, 8 data properties, 544 instances, and actively uses 5 annotation properties. The distribution between native and foreign classes and properties implemented in *HOLY* is listed in **Table 1**.

Table 1. Distribution of native and foreign classes and properties in *HOLY*.

Class and Property Type	Main Classes	Sub-classes	Object Properties	Data Properties	Annotation Properties
Native	4	97	18	0	0
Foreign	2	6	17	8	5
Total	6	103	35	8	5

3.4 Evaluation

An evaluation was performed following various criteria. In order to evaluate the RDF implementation of the *Hydrogen Ontology*, it was cross-checked for ontology pitfalls. These ontology pitfalls cover human understanding, logical consistency, elimination of modeling issues and errors in the representation of the real world, and compliance with the ontology language [36]. The validation was realized with the *OntOlogy Pitfall Scanner* (*OOPS!*). *OOPS!* is a web-based ontology checker which helps to validate ontologies by identifying previously-defined ontology pitfalls. Evaluation results of *OOPS!* contain a list of pitfalls and their severity alongside the categories 'minor', 'important', and 'critical' [37].

When cross-checking the *Hydrogen Ontology*, 12 minor issues and one important issue were detected. Two minor issues referred to classes merging different concepts and ten referred to inverse relations not being explicitly declared. Six out of the ten inverse relations labeled as not declared are object properties directly taken from *The Organization Ontology* and the *Registered Organization Vocabulary*. Creating inverse relationships for the remaining four inverse relationships would be redundant and leads to an increase of important pitfalls of recursive definitions. The two minor issues related to classes merging different concepts come from organization types created by following NACE. Modifying the structure would hinder adherence to the NACE standard. The important issue (recursive definition) is derived from using a class and its definition from *Schema.org*. Nevertheless, the ontology's consistency and reasoning are not affected by these pitfalls, as they are caused by the generality of *OOPS!* which is necessary to cover different knowledge models.

Additionally, following the *LOT* framework, test cases were used to verify the fit to functional requirements. Throughout the evaluation, four approaches using SPARQL were followed, similar to those followed by Lisena et al. [38].

1. **Explicit Relations:** query for relations that directly connect classes
   Example: ‘Which applications are there for a given product?’
2. **Inference/Aggregation:** query using aggregation or inference (e.g., group by, count, property chains, etc.)
   Example: ‘In which vehicles are PEM fuel cells more often used?’
3. **Linked Open Data:** query required information outside the scope of the model, but accessible through Linked Open Data (e.g., DBpedia)
   Example: ‘In which countries is a given company present?’
4. **OBIE:** query requiring an extension of the model or information extraction through NLP, as is present in an OBIE system.
   Example: ‘Do product components change over time?’

Some test cases required additional information obtainable through an OBIE system. As an example, assume that we are interested in the change of PEMFC components over time. In that case, we require a temporal structure and data gathered across a time span. Hence, in the fourth approach, information from the Atlant-H use case was used to evaluate whether requirements could be answered; we consider all other cases to be satisfied by the ontology. **Table 2** shows a summary of our results classified by test case type.

Table 2. Type of SPARQL Queries by Test Case type for *HOLY* listing.

Type	Explicit Relation	Inference/Aggregation	LOD	OBIE	Total
Technological	5	2	0	1	8
Market	3	1	6	0	10
Total	8	3	6	1	18

3.5 Publishing

The *Hydrogen Ontology* is implemented in RDF format and published[21] under a Creative Commons 4.0 CC-BY-SA license. The ontology requirements specification process as well as validation and evaluation results are available in the resource repository[22]. Ontology documentation is accessible online[23]. Other related resources developed and published in the context of this work are attainable through the Future Engineering homepage[24].

3.6 Maintenance

In order to support the continued use of the ontology, a maintenance plan has been developed which aims to continually evaluate the structure and accuracy of the ontology and provide feedback regarding areas of high potential for future growth endeavors. For bug detection, we employ the GitHub issue tracker which keeps control of the list of issues. For the identification of new requirements, a questionnaire[25] consisting of fixed, mostly open-ended questions has been made publicly available to allow for external input about the current state of the hydrogen market. Submissions will be regularly reviewed and considered regarding changes for further versions of *HOLY*.

4 Contribution and Future Work

In this paper, we introduced the *Hydrogen Ontology*, a domain ontology modeling the hydrogen economy. We followed the *LOT* framework and developed an ontology which combines established business models with technological domain knowledge. As such, *HOLY*'s structure is composed of two orthogonal dimensions. The first comprises six main classes representing market structures while the second organizes technological knowledge in hierarchical structures to enable the classification of products

---

21 HOLY URL: https://purl.org/holy/ns
22 HOLY repository: https://purl.org/holy/repository
23 HOLY Documentation: https://purl.org/holy/doc
24 HOLY Website: https://purl.org/holy
25 Maintenance feedback questionnaire: https://purl.org/holy/feedback

along the hydrogen value chain and the identification of components and substitutes. We included existing ontologies and vocabularies to address reusability challenges and allow expansion within other segments. To handle heterogeneous and fast-growing data sources, the model includes a lexical-semantic layer, which provides the necessary information to aid NLP of texts in the hydrogen domain, thus facilitating the construction of an OBIE system around the ontology. Additionally, the ontology was validated through a pitfall scanner and evaluated to ensure the satisfaction of its functional requirements.

At the time of writing, the published version of the ontology is being used in the Atlant-H project and is planned to be applied in the follow-up project from Atlant-H, also in cooperation with the Fraunhofer IIS. As part of this project, we intend to expand and improve the *HOLY* model. Thus, future research may extend the conceptual model to other hydrogen technologies or value chain stages. Similarly, classes like *Projects* and *Indicators* can be further detailed (e.g., via subclasses) to provide a more comprehensive market representation. Moreover, to ensure support of OBIE systems, integrated lexical-semantic information for NLP should be further evaluated to ensure proper coverage concerning available natural text in the domain (e.g., hydrogen market-related press releases, news, and publications). Furthermore, *HOLY's* applicability is not limited to the Atlant-H project alone, as it is intended to be employed in other third-party hydrogen projects such as the DuraFuelCell project, a German national research project led by the H2Ohm.

Resource Availability Statement: Source code for *HOLY*, test cases, ontology requirements specification, use case specification, maintenance feedback questionnaire, and Atlant-H use case are available from Github[26]. The *Hydrogen Ontology* and its documentation are available from Zenodo[27]. The *HOLY* website is available from the Technische Hochschule Nürnberg Georg Simon Ohm[28]. The *Hydrogen Ontology* is published under the Creative Commons 4.0 CC-BY-SA license.

Acknowledgements. This work has been supported by the Future Engineering Research Group, a research collaboration between the Technische Hochschule Nürnberg Georg Simon Ohm and the Fraunhofer IIS. In addition, we would like to thank Rene Dorsch from the Fraunhofer IIS for the useful discussion and comments on the manuscript.

Author Contribution. Conceptualization, K.A.; Methodology, K.A., C.N.; Software, K.A., C.N.; Validation, K.A.; Formal Analysis, C.N.; Investigation, K.A., C.N.; Data Curation, K.A., C.N., K.W.; Writing-original draft, K.A., C.N.; Writing-review and editing, K.A., C.N., K.W., R.Z., R.B.; Visualization, C.N., K.W.; Supervision, K.A., R.Z., R.B.; Project Administration, K.A. All authors have read and approved the final version of the manuscript.

[26] GitHub: https://purl.org/holy/repository
[27] Zenodo: https://doi.org/10.5281/zenodo.7447958
[28] HOLY Website: https://purl.org/holy

References

1. Global Market Insights: Hydrogen Market Size, Forecast Report 2021-2027, https://www.gminsights.com/industry-analysis/hydrogen-market, last accessed 2022/08/03.
2. IRENA: Geopolitics of the Energy Transformation: The Hydrogen Factor. International Renewable Energy Agency, Abu Dhabi (2022).
3. McKinsey & Company, Hydrogen Council: Hydrogen Insights 2022. Hydrogen Council (2022).
4. Fan, L., Tu, Z., Chan, S.H.: Recent development of hydrogen and fuel cell technologies: A review. Energy Rep. 7, 8421–8446 (2021). doi: 10.1016/j.egyr.2021.08.003.
5. Zuo, L., Salvadores, M., Imtiaz, S.H., Darlington, J., Gibbins, N., Shadbolt, N.R., Dobree, J.: Supporting Multi-view User Ontology to Understand Company Value Chains. In: Bernstein, A., Karger, D.R., Heath, T., Feigenbaum, L., Maynard, D., Motta, E., and Thirunarayan, K. (eds.) The Semantic Web - ISWC 2009. pp. 925–940. Springer Berlin Heidelberg, Berlin, Heidelberg (2009). doi: 10.1007/978-3-642-04930-9_58.
6. El-Gohary, N.M., El-Diraby, T.E.: Domain Ontology for Processes in Infrastructure and Construction. J. Constr. Eng. Manag. 136, 730–744 (2010). doi: 10.1061/(ASCE)CO.1943-7862.0000178.
7. Haghgoo, M., Sychev, I., Monti, A., Fitzek, F.: SARGON – Smart energy domain ontology. Smart Cities. (2020). doi: 10.1049/iet-smc.2020.0049.
8. Leontis, N.B., Altman, R.B., Berman, H.M., Brenner, S.E., Brown, J.W., Engelke, D.R., Harvey, S.C., Holbrook, S.R., Jossinet, F., Lewis, S.E., Major, F., Mathews, D.H., Richardson, J.S., Williamson, J.R., Westhof, E.: The RNA Ontology Consortium: An open invitation to the RNA community. RNA. 12, 533–541 (2006). doi: 10.1261/rna.2343206.
9. Poveda-Villalón, M., Fernández-Izquierdo, A., Fernández-López, M., García-Castro, R.: LOT: An industrial oriented ontology engineering framework. Eng. Appl. Artif. Intell. 111, 104755 (2022). doi: 10.1016/j.engappai.2022.104755.
10. Nguyen, Q.-D., Roussey, C., Poveda-Villalón, M., de Vaulx, C., Chanet, J.-P.: Development Experience of a Context-Aware System for Smart Irrigation Using CASO and IRRIG Ontologies. Appl. Sci. 10, 1803 (2020). doi: 10.3390/app10051803.
11. Lionel Tailhardat, Yoan Chabot, RaphaÃnl Troncy: NORIA-O: An Ontology for Anomaly Detection and Incident Management in ICT Systems. IOS Press. 13 (2022).
12. Espinoza-Arias, P., Poveda-Villalón, M., Corcho, O.: Using LOT methodology to develop a noise pollution ontology: a Spanish use case. J. Ambient Intell. Humaniz. Comput. 11, 4557–4568 (2020). doi: 10.1007/s12652-019-01561-2.
13. Mike de Roode, Alba Fernández-Izquierdo b, Laura Daniele, María Poveda-Villalón, Raúl García-Castro: SAREF4INMA: a SAREF extension for the Industry and Manufacturing domain. IOS Press. 17.
14. Stjernholm, M., Poslad, S., Zuo, L., Sortkj, O., Huang, X.: The EDEN-IW Ontology Model for Sharing Knowledge and Water Quality Data between Heterogeneous Databases. In: EnviroInfo2004.
15. Munir, K., Sheraz Anjum, M.: The use of ontologies for effective knowledge modelling and information retrieval. Appl. Comput. Inform. 14, 116–126 (2018). doi: 10.1016/j.aci.2017.07.003.

16. Seurin, P., Olabanjo, O., Wiggins, J., Pratt, L., Rana, L., Yasaei, R., Renard, G.: H2-Golden-Retriever: Methodology and Tool for an Evidence-Based Hydrogen Research Grantsmanship, http://arxiv.org/abs/2211.08614, (2022). doi: 10.48550/arXiv.2211.08614.
17. Lin, R., Fang, Q., Wu, B.: Hydro-graph: A knowledge graph for hydrogen research trends and relations. Int. J. Hydrog. Energy. 45, 20408–20418 (2020). doi: 10.1016/j.ijhydene.2019.12.036.
18. Zhang, M., Yang, R., Xu, J.: Construction and Analysis of Scientific Research Knowledge Graph in the Field of Hydrogen Energy Technology. In: Liang, Q., Wang, W., Mu, J., Liu, X., and Na, Z. (eds.) Artificial Intelligence in China. pp. 392–402. Springer Nature, Singapore (2023). doi: 10.1007/978-981-99-1256-8_46.
19. European Commission. Joint Research Centre.: Text mining for horizon scanning: an insight into agricultural research and innovation in Africa. Publications Office, LU (2020).
20. Moro, A., Joanny, G., Moretti, C.: Emerging technologies in the renewable energy sector: A comparison of expert review with a text mining software. Futures. 117, 102511 (2020). doi: 10.1016/j.futures.2020.102511.
21. European Commission. Joint Research Centre.: Weak signals in science and technologies in 2021: technologies at a very early stage of development that could impact the future. Publications Office, LU (2022).
22. World Wide Web Consortium: The Organization Ontology, http://www.w3.org/TR/2014/REC-vocab-org-20140116/, (2014).
23. World Wide Web Consortium: Registered Organization Vocabulary, http://www.w3.org/TR/2013/NOTE-vocab-regorg-20130801/, (2013).
24. EUROSTAT: NACE rev. 2. Office for Official Publications of the European Communities, Luxembourg (2008).
25. Lindsay Prosser ed: UK standard industrial classification of economic activities 2007 (SIC 2007): structure and explanatory notes. Palgrave Macmillan, Basingstoke, Hampshire, (2008).
26. United Nations ed: International Standard industrial classification of all economic activities (ISIC). United Nations, New York (2008).
27. Roman, D., Alexiev, V., Paniagua, J., Elvesæter, B., von Zernichow, B.M., Soylu, A., Simeonov, B., Taggart, C.: The euBusinessGraph ontology: A lightweight ontology for harmonizing basic company information. Semantic Web. 13, 41–68 (2021). doi: 10.3233/SW-210424.
28. Saggion, H., Funk, A., Maynard, D., Bontcheva, K.: Ontology-Based Information Extraction for Business Intelligence. In: Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., and Cudré-Mauroux, P. (eds.) The Semantic Web. pp. 843–856. Springer Berlin Heidelberg, Berlin, Heidelberg (2007). doi: 10.1007/978-3-540-76298-0_61.
29. Rizvi, S.T.R., Mercier, D., Agne, S., Erkel, S., Dengel, A., Ahmed, S.: Ontology-based Information Extraction from Technical Documents: In: Proceedings of the 10th International Conference on Agents and Artificial Intelligence. pp. 493–500. SCITEPRESS - Science and Technology Publications, Funchal, Madeira, Portugal (2018). doi: 10.5220/0006596604930500.
30. Noy, N.F., McGuinness, D.L.: Ontology Development 101: A Guide to Creating Your First Ontology.

31. Suárez-Figueroa, M.C., Gómez-Pérez, A., Motta, E., Gangemi, A. eds: Ontology Engineering in a Networked World. Springer Berlin Heidelberg, Berlin, Heidelberg (2012). doi: 10.1007/978-3-642-24794-1.
32. Porter, M.E.: Towards a dynamic theory of strategy. Strateg. Manag. J. 12, 95–117 (1991). doi: 10.1002/smj.4250121008.
33. Peroni, S.: A Simplified Agile Methodology for Ontology Development. In: Dragoni, M., Poveda-Villalón, M., and Jimenez-Ruiz, E. (eds.) OWL: Experiences and Directions – Reasoner Evaluation. pp. 55–69. Springer International Publishing, Cham (2017). doi: 10.1007/978-3-319-54627-8_5.
34. Konstantin Schekotihin, Patrick Rodler, Wolfgang Schmid, Matthew Horridge, Tania Tudorache: Test-driven Ontology Development in Protégé. Proc. 9th Int. Conf. Biol. Ontol. ICBO 2018. (2018).
35. Rudnicki, R., Smith, B., Malyuta, T., Mandrick, W.: Best Practices of Ontology Development. 16 (2016).
36. Lehmann, J., Isele, R., Jakob, M., Jentzsch, A., Kontokostas, D., Mendes, P.N., Hellmann, S., Morsey, M., van Kleef, P., Auer, S., Bizer, C.: DBpedia – A large-scale, multilingual knowledge base extracted from Wikipedia. Semantic Web. 6, 167–195 (2015). doi: 10.3233/SW-140134.
37. Poveda-Villalón, M., Suárez-Figueroa, M.C., Gómez-Pérez, A.: Validating Ontologies with OOPS! In: ten Teije, A., Völker, J., Handschuh, S., Stuckenschmidt, H., d'Acquin, M., Nikolov, A., Aussenac-Gilles, N., and Hernandez, N. (eds.) Knowledge Engineering and Knowledge Management. pp. 267–281. Springer Berlin Heidelberg, Berlin, Heidelberg (2012). doi: 10.1007/978-3-642-33876-2_24.
38. Lisena, P., Schwabe, D., van Erp, M., Troncy, R., Tullett, W., Leemans, I., Marx, L., Ehrich, S.C.: Capturing the Semantics of Smell: The Odeuropa Data Model for Olfactory Heritage Information. In: Groth, P., Vidal, M.-E., Suchanek, F., Szekley, P., Kapanipathi, P., Pesquita, C., Skaf-Molli, H., and Tamper, M. (eds.) The Semantic Web. pp. 387–405. Springer International Publishing, Cham (2022). doi: 10.1007/978-3-031-06981-9_23.