An Ontology Design Pattern for Industry Classification in the Facilities and Industries Ontology (FIO)

Katrina Schweikert[1,*], Torsten Hahmann[1]

[1]School of Computing and Information Science, University of Maine, Orono, ME 04469, USA

Abstract

Facilities are specific locations where commercial or institutional activity occurs. Industry classification schemes divide economic activities into functional groups. These two key concepts are crucial for analyzing institutional and economic activity and its spatial distribution, which are critical for understanding environmental impact, labor dynamics, supply chains, and many other complex social and environmental phenomena. In this paper, we present the Facilities and Industries Ontology (FIO) as a generic ontology design pattern for linking facilities to industry sectors as defined in classification systems. FIO supports semantic reasoning to, e.g., infer broader industry sectors that are associated with facilities and organizations. We evaluate the pattern by constructing and querying an ontology-based knowledge graph of the facilities in the continental U.S. as recorded in the US EPA Facility Registry Service (FRS) with their classifications by the North American Industry Classification System (NAICS).

Keywords

ontology, ontology design pattern, industry classification, NAICS, industrial facilities, facility registry service

1. Introduction

Facilities—such as industrial plants, restaurants or gas stations, infrastructure hubs like airports, or municipal services like landfills or wastewater treatment plants—play a central role in how societies function. They serve as the operational sites where goods are produced, services are delivered, and resources are managed. To organize and study facilities more systematically, industry classification, like the North American Industry Classification System (NAICS) [1], have been developed. Together, facilities and their associated industries provide essential information for understanding and addressing major societal challenges. From environmental sustainability and economic development to public health and urban planning, the types, locations, and functions of facilities help shape policy decisions and community outcomes. Industries define patterns of resource consumption, pollution, employment, and demographic change. By analyzing facilities within the context of their associated industries, we can gain valuable insights into how society can respond more effectively to complex issues related to economic and environmental resilience and sustainability. An ontology that explicitly and systematically describes facilities and their categorization into industries can facilitate such analyses. This paper describes our work on developing a general pattern for such an ontology—which we refer to as the Facilities and Industries Ontology (FIO) Ontology Design Pattern (ODP)—for describing facilities and identifying the industry that they belong to. We further show how we used and refined the pattern to construct an ontology-based knowledge graph of the facilities in the continental U.S. as recorded in the Facility Registry Service (FRS) [2] maintained and shared by the U.S. Environmental Protection Agency (EPA) and their classification according to NAICS and the Standard Industry Classification (SIC) [3].

The contribution of the FIO ODP is to create a simple and general pattern for relating facilities to industries and likewise for identifying the locations of economic activity in different industries. The pattern accommodates any hierarchically structured industry classification schema and can handle multiple competing industry schemata. It reuses and integrates seamlessly with standardized ontologies

Proceedings of the Joint Ontology Workshops (JOWO) - Episode XI: The Sicilian Summer under the Etna, co-located with the 15th International Conference on Formal Ontology in Information Systems (FOIS 2025), September 8-9, 2025, Catania, Italy

*Corresponding author.

katrina.schweikert@maine.edu (K. Schweikert); torsten.hahmann@maine.edu (T. Hahmann)

0000-0003-3271-6700 (K. Schweikert); 0000-0002-5331-5052 (T. Hahmann)

for provenance and geospatial knowledge (i.e. PROV-O [4] and Geosparql [5]). We provide an OWL2 implementation of FIO, that leverages semantic inferencing to simplify querying facilities by industry and finding where particular industries are active. We extend the pattern and populate it with data from FRS and NAICS to demonstrate and evaluate its use. We also present reusable and minted IRIs for these widely referenced entities in the FRS and NAICS datasets.

2. Use Case

The development of the FIO pattern and its implementation as a Knowledge Graph are motivated primarily by the needs of the SAWGraph project [6]. This use case and select motivating competency questions are described next before we discuss the pattern's broader applicability across other domains and use cases in Section 2.3.

2.1. Primary Use Case: Tracing Environmental Contamination by PFAS

The development of FIO is part of a larger effort—the SAWGraph project [6] that is part of NSF's Proto-OKN [7]—to build an Open Knowledge Network to support understanding and analyzing environmental contamination by Per and Poly-FlouroAlkyl Substances (PFAS). PFAS are synthetic chemicals used for their oil-, water-, and fire-resistant properties in products ranging from food packaging and stain-resistant fabrics to electronics and firefighting foams. Due to their persistence, many PFAS accumulate in organisms and magnify up the food chain, posing significant health risks to humans [8]. They have been identified as an emergent chemical of concern by the EPA and similar agencies in many other countries. PFAS do not occur naturally but are produced by chemical manufacturers. However their production, usage and release as waste is minimally tracked, and environmental testing is costly and limited to a small subset of the over 14,000 known PFAS compounds.

SAWGraph is primarily developed to help U.S. state agencies prioritize locations to test for PFAS, assess contamination distribution and impacts, and investigate potential transportation pathways from suspected point sources to accumulation sites [9]. EPA's FRS [2] database captures the facilities that release PFAS into the air or water, making knowledge of these facilities and their industrial classifications essential for understanding transportation pathways, assessing industry-specific risks, and guiding future testing. Several studies, e.g. [10, 11], have identified *industries of concern* that potentially are or were producers or users of PFAS. These industries, classified using codes from NAICS, range from PFAS-using manufacturers to airports, military bases, or firefighting training sites were substances containing PFAS may have been used extensively. PFAS also accumulates at facilities such as landfills and waste water treatment plants, from where it can spread into the surrounding water, soil or air—underscoring the importance of identifying such facilities for studying environmental contamination.

2.2. Competency Questions from the Primary Use Case

As part of the SAWGraph project, we worked with domain experts from various federal and state agencies—primarily environmental protection agencies such as the US EPA and Maine's Department of Environmental Protection—as well as academia to gather a broad set of questions. While many of these questions go beyond information about facilities and industries, they all include key components relevant to the scope of the Facility and Industry Ontology (FIO). From these broader questions we identified five common types of competency questions (CQs)—listed as items 1 through 5 below—that concern facilities and their industry classifications, and that FIO should directly support. These CQs have guided FIO's development. For each of them, we also provide examples of the broader SAWGraph questions that motivate them underneath. Answering these broader questions requires linking FIO to other ontologies and graphs still under development.

1. Retrieve all **facilities** of **NAICS code 562212** ***located in*** Penobscot County, Maine.
    - Retrieve all PFAS samples near waste collection facilities (NAICS 5622) in Maine.

- Retrieve all landfills in Maine that are near any waterbody.

2. What **NAICS industry subsector** is **Penobscot Energy Recovery facility** (FRS id 11000991341) ***associated with***?
   - Is this facility in an industry suspected of handling PFAS?
   - How many facilities of industries that are suspected of handling PFAS are within 1km of the Kennebec River?
3. Find all **facilities *located in*** a given set of S2 cells.
   - Which facilities are upstream of samplepoints with a reported PFOA concentration ≥20ppt?
   - What facilities are close to a private water supply wells with a concentration of at least 20ppt of any PFAS?
4. Retrieve all **facilities *located in*** Maine in the **NAICS Industry Groups 3221** (Pulp, Paper, and Paperboard Mills) and **3222** (Converted Paper Product Manufacturing), their specific **NAICS Industry Codes** and the county they are ***located in***.
   - How many facilities of industries that are suspected of handling PFAS are in the surface water protection area of this public water supply?
   - Which counties in Maine have paper manufacturing facilities? What industry group are they associated with?
5. What **subsectors** of the **"Manufacturing" sector** (NAICS codes 31-33) are associated with facilities suspected of handling PFAS, and which of them have facilities ***located in*** Maine?
   - Which subsectors of facilities are near high PFAS test results or have recorded PFAS releases across the country? Which facilities of those subsectors exist in this state that has had limited testing and reporting of PFAS?

2.3. Broader Use in Other Domains

The proposed pattern and its extension as well as their implementation in the form of a KG have a much broader application potential. It is of immediate wider use across environmental monitoring to analyze who, where and what chemical and biological pollutants are emitted into the air, water or soil. This covers inorganic chemicals such as heavy metals (e.g. arsenic or mercury), bacteria and viruses, as well as emissions of gases (e.g. carbon monoxide and dioxide, sulfur dioxide, or ozone). Moreover, it can be used to help with resource and supply chain management, e.g., to determine where certain kinds of raw materials or expertise may be needed or where energy-intensive industries are concentrated. Likewise, correlating the locations of facilities of certain industries to demographic data can reveal patterns of environmental burden on vulnerable populations. At the same time, the knowledge of facilities and industries can help forecast where growth in terms of employment, revenue, and the need for housing, transportation, electricity generation, or water and wastewater treatment may be concentrated. It can be used to shape workforce and regional development plans and direct infrastructure investments. Likewise, the information can help detect industrial and economic diversification and assess regional resilience to economic shocks. Some of these topics are addressed by other Proto-OKN projects (see https://www.proto-okn.net/theme-1-projects/ for their descriptions) and it is expected that the FIO KG will serve as a shared resources available to multiple ongoing and future data and knowledge integration projects.

3. Related Work

FIO draws and links widely used concepts—*facility*, its *location*, and associated *organization*—that appear in some variant in many existing ontologies, such as GoodRelation [12] , schema.org [13], the DataCommons ontology [14], the Organization Ontology [15], the Financial Industry Business Ontology (FIBO) [16], and the SAREF suite of ontologies (https://saref.etsi.org/) . However, these related ontologies differ from FIO in the overall scope, the interpretation of the key concepts, the absence of a distinct "industry" concept, and the extent of formalization.

Different scopes The majority of related ontologies are designed for specific domains, which limits their generality. For example, FIBO is focused specifically on providing a terminology for the financial industry, which includes terms like `NAICS code` and `SIC code`. However, because organizations (e.g. companies) are the key entities of interest to finance use cases (e.g. for risk or profit analysis), it associates NAICS and SIC codes exclusively with legal organizations. This does not align well with environmental, regulatory, demographic, or land use planning use cases where facilities are the key entities of interest. GoodRelations contains relevant concepts like `BusinessEntity` and `Location`, which are closely related to FIO's concepts *Organization* and *Facility*. However, its narrower focus on e-commerce means it primarily represents the products and services offered by organizations. Schema.org reuses and builds on the concepts and properties from GoodRelations albeit with differences in names, e.g. `Place` instead of `Location`. The Organization Ontology is focused on modeling organizations as legal entities, along with relationships like subsidiaries or parent organizations. It includes classes such as `Organization`, `Site`, and `OrganizationalUnit`, but does not model industries explicitly. Some of these concepts are also reused in SAREF for the IoT domain, though its concepts are reused from the Organization ontology or used in much narrower contexts.

Different interpretation of the key *facility* concept Even when ontologies include similar terms or concepts, their interpretation differs in drastic or more subtle ways. For example, FIBO defines a `Facility` based on the capabilities it provides to an organization, and permits *virtual facilities* as valid instances. Likewise, the Organization ontology `Site` class and schema.org's `Location` class are somewhat similar to a *Facility* in FIO, but also allow sites to be virtual. In contrast, FIO focuses exclusively on facilities as *physical sites*.

Absence of an explicit *industry* concept and its semantics A major gap in the surveyed ontologies is the lack of an explicit *industry* concept in all of them but FIBO. While GoodRelations and schema.org allow associating industry classification codes with organizations using dataype properties—`hasNAICS` and `hasISICv4` in GoodRelations, and `naics` and `isicv4` in schema.org—these properties point to literal values rather than entities in their own right (i.e., classes or instances representing specific industries or industry sectors). As a result, industry codes are not semantically represented and cannot be meaningfully related to one another. Datacommons, which builds on schema.org, offers linked data about facilities, sites of facilities and organizations in the U.S. from the Bureau of Labor Statistics, the Census Bureau, and the EPA. However, like schema.org itself, NAICS codes are represented as unstructured literals. This again precludes the ability to connect industries hierarchically (e.g. the fact that the "Beverage Manufacturing" industry—NAICS code 3121—is a subsector of the broader "Manufacturing" industry—NAICS codes 31–33) or to capture other semantics of industries more formally. While FIBO does provide *industry sector classifier* as an explicit concept with subclasses for NAICS, SIC and other industry codes, it too lacks semantic relationships between the classifiers.

More broadly, none of these ontologies support the kind of reasoning required to answer the motivating competency questions— for example retrieving all subsectors of an industry sector or all facilities within a broader industry sector due to a lack of explicit semantics about industries and their associations with facilities, organizations, and other industries.

FIO's complementary nature While existing ontologies do not fully meet the specific requirements that motivated the development of FIO and were illustrated by the CQs—such as reasoning over facilities and industries at varying levels of granularity—FIO is intended to complement them. It functions as what can be described as a generic reference pattern: an ontology design pattern that provides a high-level unified view across ontologies within a single or across closely related domains, in the spirit of a domain reference ontology [17] or, similarly, a domain-related ontology pattern [18]. As such, more comprehensive domain ontologies can be treated as extensions of FIO to facilitate their alignment and enhance their utility in cross-domain and federated data applications. FIO's abstract yet tightly integrated model of facilities, industries, and organizations serves as a bridging layer to support interoperability and semantic integration across these diverse ontologies. For example, integration with FIBO would allow enrichment with financial, legal, and organizational attributes, while alignment with the Organization Ontology would support more detailed modeling of organizational structures, memberships, and roles. Similarly, FIO would complement other domain-specific ontologies, such as the COGITO Facility Ontology [19] that models detailed physical structures and infrastructure components of facilities in the construction domain.

4. Approach

To develop FIO, we adopt a bottom-up approach grounded in two artifacts: (1) the set of *competency questions* from SAWGraph's environmental contamination use case presented in Section 2.2, and (2) the *FRS dataset* [2] that provides the data to populate the knowledge graph and supports evaluation. Based on the CQs, we design the core of FIO to capture essential concepts, relationships, and attributes associated with facilities and industries (Section 5) and implement it as a reusable OWL 2 ontology (Section 5.4). We then develop two dataset-specific extensions: one for modeling the NAICS industry classification scheme (Section 6.2) and one for EPA's Facility Registry Service (FRS), the latter introduced in Section 6.1 but still under development and to be detailed in future work. After deploying FIO and its extensions in a knowledge graph and populating it with data from FRS, we evaluate it by translating the five motivating CQs into SPARQL queries and executing them on the constructed KG (Section 7).

5. FIO's Core Pattern

The core concepts of the Facilities and Industries Ontology Design Pattern (FIO) are *Facility* and *Industry*, reflecting their prominence in the competency questions. A *Facility* is defined as a physical entity with a fixed geospatial location where commercial or institutional activity occurs or has occurred in the past. An *Industry* represents a particular economic subdivision characterized by its function and services. For example, grain farming and grain milling represent distinct industries within the farming and food manufacturing sectors, respectively. Facilities and industries are closely connected: each facility is associated with one or more industries, modeled in FIO using the *ofIndustry* property. The following sections detail the modeling of facilities, industries, and their interrelationships.

5.1. Facility

Unlike other related ontologies, FIO restricts *Facility* to physical entities with fixed geospatial locations where commercial or institutional activity occurs. This design choice excludes virtual facilities or sites and helps more easily align with top-level ontologies that distinguish material from non-material entities. Accordingly, *Facility* is modeled as a subclass of GeoSPARQL's *geo:Feature* class and may represent a building, building complex, building part, or a site such as an airstrip, mine, or Superfund site. Because

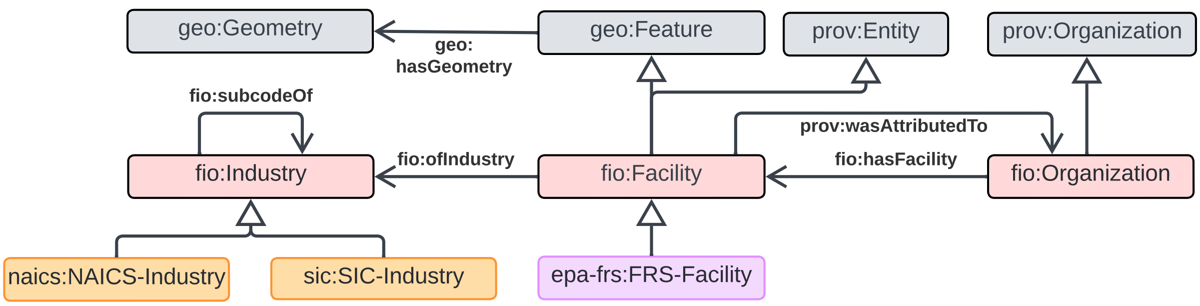

Figure 1: FIO's basic conceptual model, its integration with more generic concepts from reused ontologies, and example extensions by classes that represent FRS facilities and NAICS and SIC industries, respectively.

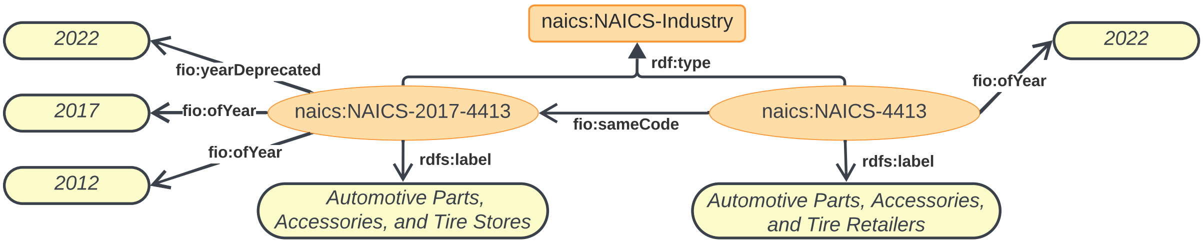

Figure 2: Modeling the changes to Industry Codes over time. Here, the NAICS code 4413 has been updated in 2022. The old code is represented by the entity `naics:NAICS-2017-4413` was included in the NAICS classification in 2021 and 2017 but was deprecating in 2022. It was superseded—as indicated by the *sameCode* property—by the new entity `naics:NAICS-4413` whose label has a minor update.

location is a defining characteristic, each *Facility* is linked to a *geo:Geometry* (e.g. a point location or polygon) via GeoSPARQL's *geo:hasGeometry* property, with optional address details provided as text using the *schema:address* property. Creating facilities as *geo:Features* also enables precomputation of spatial relations with grid cells (compare [20]) and administrative regions. In SAWGraph, for example, containment of facilities within level 13 S2 cells (of roughly $1km^2$) and level 3 administrative regions (e.g., towns or townships in the US) are precomputed and stored using the *kwg-ont:sfWithin* property as exemplified in Figure 3 and described in more detail in [9].

Another key aspect of a facility is its ownership. It is encoded using Prov-O's *prov:wasAttributedTo*, which links to the *Organization* that owns or manages the *Facility*. *Organization* specializes Prov-O's *prov:Organization* class, which makes *Facility* a subclass of *prov:Entity*. Additional facility-specific details can be represented using dataset-specific extensions as needed.

5.2. Industry

Industry classification systems categorize economic activity based on economic function and services provided. In FIO, we model these classifications by abstracting from specific classification schemes such as the North American Industry Classification System (NAICS) [1], Standard Industrial Classification System (SIC) [3], and Global Industry Classification Standard (GICS) [21]. All of them organize industries hierarchically using *classification codes*, ranging from broad sectors to highly specific industries.

FIO represents each industry—regardless of its level of specificity—using the *Industry* class. We use the term *industry code* interchangeably because it identifies the same entity—the *Industry*—despite their differing ontological natures. To model the hierarchical structure of industry codes, we introduce the transitive object property *subcodeOf* between pairs of *Industry* instances. This reflects how industry classification codes are used in practice; their specificity often depends on the data collection purpose and context. Queries also often target higher-level categories such as industry groups or subsectors.

Because industry classification schemes evolve over time to reflect emerging sectors, we add the properties *ofYear* and *yearDeprecated* to indicate when an industry was added or removed from a classification scheme[1]. To avoid unnecessary duplication, we do not create new instances for all industries for each new version of a scheme. Instead, we attach all applicable years to a single *Industry* instance and only create new instances when anything changes. Figure 2 examplifies this approach.

5.3. Facility-Industry Associations

To represent the association of facilities with industries, we introduce the *ofIndustry* object property. It is intended to link a *Facility* to one or more *Industry* instances at any level of specificity and from one or multiple classification schemes—even beyond NAICS or SIC. It accommodates differences in the nature of the facility's activities and the granularity of the available data. The assignment of industries to a facility can depend on several factors: the classification scheme in use, the granularity of its categories,

[1]For example, NAICS is updated every five years; schemes generally change no more than once per year.

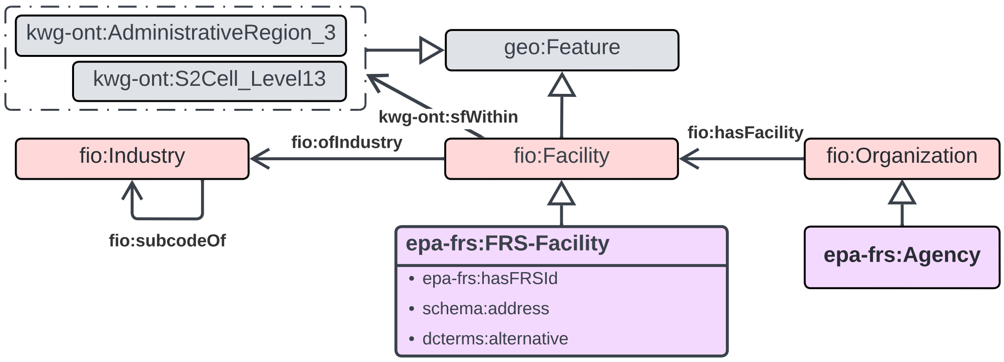

Figure 3: The dataset-specific extension of FIO for representing FRS Facilities and their location relative to spatial reference entities, such as administrative regions or cells from the S2 reference grid.

and the diversity and complexity of activities occurring at a given facility. A facility may thus be linked to one or multiple industries, potentially spanning different classification systems.

In many public datasets, however, industry codes are assigned not to facilities but to organizations. FIO permits this usage as well: the domain of the *ofIndustry* property includes both *Facility* and *Organization*. Nonetheless, FIO's pattern and axiomization are specifically designed to support answering spatial questions about *where* industry-related activities take place, which requires the use of facilities.

To support basic reasoning over the hierarchical nature of industry classification schemes, FIO includes a property chain axiom that allows inferring new, implicit instances of *subcodeOf* from any composition of *ofIndustry* and *subCodeOf*, thereby indirectly associating a facility with all broader industry categories it belongs to. As a result, users can easily query for facilities using any level of industry categories regardless of the specificity of their directly assigned industries.

5.4. OWL2 Implementation

FIO is implemented in OWL2 [22] using Turtle syntax and is publicly available at https://github.com/SAWGraph/fio. The use of permanent identifiers in the subdomain w3id.org/fio is planned. The OWL2 implementation has been validated using the Pellet Reasoner as executed from the Protégé software, and Protégé has been used to inspect its coherency and consistency [23]. The FIO pattern itself (without the extensions) has minimal ontological commitment, it newly defines only three classes and six properties, while reusing classes and properties from the GeoSPARQL, PROV-O and DCTERMS ontologies. The use of property chain axioms requires usage of the OWL2 RL profile to support the full intended inferencing.

6. Extending and Populating the FIO Pattern for Use in a KG

To evaluate FIO, we extend it to model the dataset-specific concepts and relations from EPA's Facility Registry Service (FRS) [2] and the associated industries from the NAICS classification scheme [1]. This supports the SAWGraph primary use case, and anticipates reuse by other Proto-OKN projects.

Each dataset-specific extension of FIO uses its own namespace. This allows for the preservation of properties and provenance unique to each dataset and supports individual reuse of the dataset-specific ontologies for other applications even without FIO.

6.1. EPA Facilities

The EPA Facility Registry Service (FRS) is a centrally managed registry that identifies facilities that are subject to environmental regulations or are of environmental interest [2]. This includes facilities subject to regulation under clean air, water or waste management regulations, including during construction. It also includes facilities that apply for environmental assistance and support programs (e.g. for

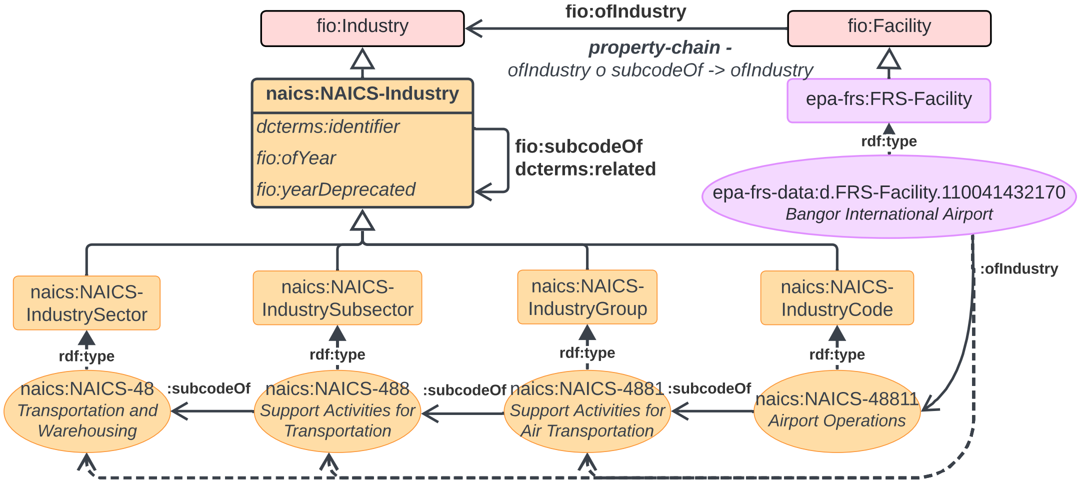

Figure 4: The dataset-specific extension of FIO for NAICS with its subclasses and example instances (oval shapes). To support queries about the classification hierarchy, we leverage the transitive *fio:subcodeOf* relation that links each instance *Industry Code* to its *Industry Group*, *Industry Subsector*, and *Industry Sector* as exemplified for the instances at the bottom.

remediation) and registration programs at the state or national level (e.g. underground storage tanks, aquifer protection programs). The FRS aligns facility information across a variety of state and federal information systems, to provide a standard facility identifier, name, and location.

The *epa-frs:FRS-Facility* class represents these facilities. Relations from de-facto ontology standards, including *dcterms:identifier*, *rdfs:label*, *dcterms:alternative*, *geo:hasGeometry*, and *schema:address*, are used to provide more detailed semantic descriptions of the facilities. We also mint permanent identifiers for all the facilities in the continental U.S. included in the FRS. Additional details on the dataset ontology and instantiation are available on GitHub at https://github.com/SAWGraph/fio and will be elaborated on in future work.

6.2. NAICS Industry and Industry Classification

NAICS [1] is an industry classification scheme for North America jointly developed by the U.S., Canada, and Mexico. Unlike with GICS, there is no governing body that assigns and catalogs industry categories for each facility or organization. In the U.S., a variety of government agencies record NAICS and/or SIC[2] codes associated with facilities or organizations depending on the purpose of the data collection. As such, one facility can be described as having multiple industry categories relating to various aspects of activities that occur there and the interest (e.g. contamination, demographics, economic output) of the data collection. Sometimes, the assigned categories also vary in their level of specificity within the industry hierarchy. To model NAICS industry codes, we create *naics:NAICS-Industry* as a single subclass of *fio:Industry* that is specific to the NAICS classification system and that serves as common superclass for all NAICS industry categories. For each level of specificity—which is reflected in NAICS in the length of the assigned industry code—we create a subclass. For example *naics:NAICS-IndustrySector* is the class that represents the coarsest industries, which are identified using two-digit codes, while *naics:NAICS-IndustryCode* represent the most fine-grained industries identified by five or six digit codes. The classes are instantiated using specific industry categories, e.g. *naics:NAICS-48* is an instance of *naics:IndustrySector*, as illustrated in Figure 4.

We instantiate the latest NAICS version (from 2022) and add separate instances for prior year codes that have content differences. They are related to the 2022 instances using the *sameCode* property. All codes that have remained unchanged over time have a *ofYear* property value for each of the years NAICS was published—from 1997 to 2022— associated with them. Only codes that have been deprecated since (e.g. by a change in meaning) have their last valid year included in the IRI.

[2]SIC [3] is a precursor of NAICS but still in use in some information systems.

7. Evaluation using the Competency Questions

We have evaluated FIO using the competency questions outlined in Section 2.2 to assess the completeness of the ontology and it's ability to meet the needs of the identified use cases. We demonstrate each competency question with a SPARQL query which uses the prefixes listed in Listing 1 . These questions were tested on a graph that includes all NAICS codes for the latest NAICS publication (2022), and all facilities in FRS in two states (Maine and New Hampshire), which constitutes a graph of more than 4 million triples, only 1.2 million of which are explicit and the remainder are inferred. It includes over 45,000 facilities with industry association from 36 different information systems that contribute data to the FRS. FIO was loaded together with the dataset ontologies and data instances for NAICS and FRS into GraphDB [24].

Listing 1: Prefixes

```
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX fio: <http://w3id.org/fio/v1/fio#>
PREFIX naics: <http://w3id.org/fio/v1/naics#>
PREFIX epa-frs: <http://w3id.org/fio/v1/epa-frs#>
PREFIX kwg-ont: <http://stko-kwg.geog.ucsb.edu/lod/ontology/>
PREFIX kwgr: <http://stko-kwg.geog.ucsb.edu/lod/resource/>
```

Competency Question 1 finds facility of a particular industry type in a specific administrative region. The ontology design pattern supports modifying this template to easily support querying facilities at any level of NAICS code, without text parsing, and any level of administrative region from county subdivision to state (i.e. administrative level 3 to administrative level 1). This query returns the seven solid waste landfill facilities in Penobscot County, Maine, listed in the result table[3]

CQ 1: Retrieve all facilities of NAICS code 562212 located in Penobscot County

```
SELECT * WHERE {
   ?facility fio:ofIndustry naics:NAICS-562212; # Solid Waste Landfill facilities
         rdfs:label ?facilityName; # facility name
         kwg-ont:sfWithin kwgr:administrativeRegion.USA.23019. } #Penobscot County(by FIPS)
```

facility	facilityName
epa-frs-data:d.FRS-Facility.110009913415	PENOBSCOT ENERGY RECOVERY FACILITY
epa-frs-data:d.FRS-Facility.110032749177	NEW ENGLAND WASTE SERVICES OF MAINE - PINE TREE LANDFILL
epa-frs-data:d.FRS-Facility.110038020049	JUNIPER RIDGE LANDFILL
epa-frs-data:d.FRS-Facility.110039664342	ORONO CDD LANDFILL
epa-frs-data:d.FRS-Facility.110040176181	TOWNWIDE MS4 STORMWATER GENERAL PERMIT
epa-frs-data:d.FRS-Facility.110055618577	ORONO WATER POLLUTION CONTROL FACILITY
epa-frs-data:d.FRS-Facility.110058407941	DOLBY LANDFILL

Question 2 tests the ability to reason about the industry hierarchy, specifically to associate a facility with a more generalized industry classification than was explicitly linked. Due to transitivity of the *subCodeOf* property, and the property chain between *ofIndustry* and *subCodeOf*, all levels of specificity of industry are inferred for the facility via *ofIndustry* relation, and the query only needs to specify which class of industry code (e.g. NAICS industry subsector) is required. This query returns one industry subsector for the specified facility, though the example facility has three NAICS industry codes associated with it, they all belong to the same subsector.

CQ 2: What NAICS industry subsector is Penobscot Energy Recovery facility associated with?

```
SELECT * WHERE {
   epa-frs-data:d.FRS-Facility.110009913415 dcterms:identifier ?id; # One facility with ID
         fio:ofIndustry ?industry. # with associated industry codes
```

[3]The full results from all competency questions are available at https://github.com/SAWGraph/fio/

```
   ?industry a naics:NAICS-IndustrySubsector; # that are NAICS Industry SubSectors
           rdfs:label ?industryLabel. }
```

Competency Question 3 demonstrates the spatial reasoning capabilities of the graph. It is derived from a number of use cases that reason about the proximity or overlap to a number of spatial features, such as waterbodies, other facilities, and sample points as discussed in [9]. This query identifies four facilities in the two specified S2 cells.

CQ 3: Find all facilities located in a given set of S2 cells.

```
SELECT * WHERE {
   ?facility kwg-ont:sfWithin ?s2. # Facilities by region
   ?s2 a kwg-ont:S2Cell_Level13. # Where region is an s2 cell
   VALUES ?s2{kwgr:s2.level13.5523882010617053184 kwgr:s2.level13.9935713923632201728}.
   } # Specific s2 cells to search
```

Competency Question 4 combines parts of previous questions that filter facilities, by both specifying a spatial region and going from general industry category to more specific industry codes. This query returns 43 facilities in 11 different counties in Maine in the specified Paper industry groups. (Note that additional filtering would be required to narrow them to currently operational facilities).

CQ 4: Retrieve all facilities located in Maine in the NAICS Industry Groups 3221 (Pulp Paper and Paperboard Mills) and 3222 (Converted Paper Product Manufacturing) and their specific NAICS Industry Codes and the county they are located in.

```
SELECT * WHERE {
VALUES ?industryGroup{naics:NAICS-3221 naics:NAICS-3222}
   ?facility a fio:Facility;
      rdfs:label ?facilityName;
      fio:ofIndustry ?industryGroup; # all facilities in the Industry Group 3221 or 3222
      fio:ofIndustry ?industryCode; # with all additional industry codes
      kwg-ont:sfWithin ?region. # by administrative region
   ?region a kwg-ont:AdministrativeRegion_3. # by county subdivision (Admin Level 3)
   ?industryCode a naics:NAICS-IndustryCode; # only NAICS specific industry codes
          rdfs:label ?industryName.
   SERVICE <repository:Spatial> #Federated query to filter region to Maine (FIPS 23)
      {?region kwg-ont:administrativePartOf+ kwgr:administrativeRegion.USA.23;
            kwg-ont:administrativePartOf ?county.
      ?county a kwg-ont:AdministrativeRegion_2; # labeled by County
            rdfs:label ?countyName.} }
```

The final competency question is built on the concept of generalizing information about specific facilities to other facilities of the same industry type. In the primary use case, due to limited tracking of PFAS chemicals and propriety information around some manufacturing processes, it is necessary to hypothesize that known chemical usage at specific facilities also likely occurs in other facilities of the same industry. This type of query also takes patterns from one region and applies it to another (a specific state). This query returns 1135 Facilities in Maine which belong to 17 different industry subsectors in manufactoring (of the 27 total manufacturing subsectors identified in NAICS).

CQ 5: What subsectors of the "Manufacturing" sector (NAICS codes 31-33) are associated with facilities suspected of handling PFAS? Which facilities of those subsectors are located in Maine?

```
SELECT * WHERE {
   #Find subsectors of manufacturing likely to use PFAS
   {SELECT DISTINCT ?industry ?industryLabel WHERE{
      ?facility a epa-frs:EPA-PFAS-Facility; # all PFAS suspected Facilities
         fio:ofIndustry ?industry.
      ?industry a naics:NAICS-IndustrySubsector; # by subsector
         rdfs:label ?industryLabel;
         fio:subcodeOf|fio:subcodeOf/owl:sameAs naics:NAICS-31. } } #in Manufacturing
```

```
  #Find all facilities in Maine in those subsectors
  ?MaineFacility a fio:Facility; # find facilities
         rdfs:label ?facilityName;
         fio:ofIndustry ?industry; # in industry returned from nested query above
              kwg-ont:sfWithin ?region.
  ?region a kwg-ont:AdministrativeRegion_3.
 SERVICE <repository:Spatial>{
  ?region kwg-ont:administrativePartOf+ kwgr:administrativeRegion.USA.23. # in Maine
    }}
```

These SPARL query translations of the Competency Questions demonstrate and validate the ability of the FIO pattern to meet the needs of the primary use cases.

8. Summary

FIO establishes relationships between the core concepts *Facility*, *Industry*, and *Organization* in a way that abstracts away domain- or task-specific details reserved for domain ontologies, while axiomatizing the semantics of these relationships as necessary for automated reasoning. In line with our vision of FIO as a generic reference pattern, we limit ontological commitments to what is needed to enable seamless querying across facilities and industries. While still in the final stages of development and in need of more in-depth evaluation, the FIO pattern and its axiomatization offer three distinct advancements:

1. **Facility as Primary Physical Actor** FIO treats facilities as the primary entities responsible (in a physical sense) for environmental impacts, such as emissions, traffic, and contamination, through their activities. Consequently, it distinguishes between *facilities* (as physical entities) and *organizations* (as legal entities that own or manage one or more facilities).
2. **Industry as Independent Semantic Concept** Rather than representing industries merely as literal values of properties of facilities or organizations, FIO represents each industry as a separate concept. This allows for relating industries via formal semantic relationships using subclass and subproperty relationships to capture the hierarchical nature of industry classifications. These can be readily used in queries.
3. **Facility-Industry Associations** FIO links facilities, rather than organizations, to industries. This supports a finer-grained specification of industrial activity. This allows more accurately pinpointing at which facilities of a large organization (e.g. vertically integrated companies like Apple or Shell or companies like Amazon or GE that are active across multiple, sometimes unrelated industries) certain industrial activities take place. For example, just because Apple is active in manufacturing (e.g. NAICS code 334111 for Computer and Peripheral Equipment Manufacturing), does not mean every one of its retail stores (NAICS code 443142) should be treated as a manufacturing facility. While one could model each facility as being one suborganization that is part of a larger organization, this risks introducing artificial organizations and blurring the ontological differences between facilities and organizations.

Acknowledgement This work and the development of SAWGraph have been supported by the National Science Foundation (NSF) under Grant No. 2333782 as part of the Proto-OKN initiative (https://www.proto-okn.net/). The views expressed are those of the authors and do not necessarily reflect those of the NSF.

Declaration on Generative AI The authors have employed ChatGPT for improving the clarity of the text. The authors reviewed and edited the content afterwards as needed and take full responsibility for the publication’s content.

References

[1] US Census Bureau, North american industry classification system, united states, 2022, https://www.census.gov/naics/reference_files_tools/2022_NAICS_Manual.pdf, 2022.

[2] US Environmental Protection Agency, Facility registry service, https://www.epa.gov/frs, 2024.

[3] US Department of Labor, Standard industry classification (sic) manual, https://www.osha.gov/data/sic-manual, 2025.

[4] T. Lebo, S. Sahoo, D. McGuinness, PROV-O: The PROV ontology, http://www.w3.org/TR/prov-o/, 2013.

[5] Open Geospatial Consortium, OGC GeoSPARQL - A Geographic Query Language for RDF Data (1.0), https://www.ogc.org/publications/standard/geosparql/, 2011.

[6] T. Hahmann, P. Hitzler, H. K. McGinty, G. Hettiarachchi, O. Apul, et al., Safe Agricultural Products and Water Graph (SAWGraph): An Open Knowledge Network to Monitor and Trace PFAS and Other Contaminants in the Nation's Food and Water Systems, https://sawgraph.github.io/, 2024.

[7] National Science Foundation, The Proto-OKN initiative, https://www.proto-okn.net/, 2023.

[8] PFAS exposure pathways for humans and wildlife: A synthesis of current knowledge and key gaps in understanding, Environ. Toxicol. Chem. 40 (2021) 631–657. doi:10.1002/etc.4935.

[9] K. Schweikert, D. Kedrowski, S. Stephen, T. Hahmann, Precomputed topological relations for integrated geospatial analysis across knowledge graphs, in: 13th International Conference on Geographic Information Science (GIScience 2025), LIPIcs 346, 2025, pp. 4:1–21.

[10] D. Salvatore, K. Mok, K. K. Garrett, G. Poudrier, P. Brown, L. S. Birnbaum, G. Goldenman, M. F. Miller, S. Patton, M. Poehlein, J. Varshavsky, A. Cordner, Presumptive contamination: A new approach to PFAS contamination based on likely sources, Environmental Science & Technology Letters 9 (2022) 983–990. doi:10.1021/acs.estlett.2c00502.

[11] US Environmental Protection Agency, Multi-Industry Per-and Polyfluoroalkyl Substances (PFAS) Study-2021 Preliminary Report, Technical Report, 2021.

[12] M. Hepp, GoodRelations: An ontology for describing products and services offers on the web, in: 16th Intern. Conf. on Knowledge Eng. and Knowledge Mgmt. (EKAW 2008), 2008, pp. 329–346.

[13] R. V. Guha, D. Brickley, S. Macbeth, Schema.org: evolution of structured data on the web, Commun. ACM 59 (2016) 44–51. doi:10.1145/2844544.

[14] R. V. Guha, P. Radhakrishnan, B. Xu, W. Sun, C. Au, A. Tirumali, M. J. Amjad, S. Piekos, N. Diaz, J. Chen, J. Wu, P. Ramaswami, J. Manyika, Data commons, 2023. arXiv:2309.13054.

[15] D. Reynolds, The organization ontology, https://www.w3.org/TR/vocab-org/, 2014.

[16] M. Bennett, The financial industry business ontology: Best practice for big data, Journal of Banking Regulation 14 (2013) 255–268.

[17] T. Hahmann, S. Stephen, Using a hydro-reference ontology to provide improved computer-interpretable semantics for the groundwater markup language (GWML2), International Journal of Geographic Information Science 32 (2018) 1138–1171. doi:10.1080/13658816.2018.1443751.

[18] R. D. A. Falbo, G. Guizzardi, A. Gangemi, V. Presutti, Ontology patterns: clarifying concepts and terminology., in: 4th Workshop on Ontology and Semantic Web Patterns (WOP 2013), 2013. URL: https://ceur-ws.org/Vol-1188/paper_11.pdf.

[19] M. P. Villalón, R. G. Castro, S. B. Galindo, Facility ontology, https://cogito.iot.linkeddata.es/def/facility/, 2025.

[20] S. Stephen, M. Faulk, K. Janowicz, C. Fisher, T. Thelen, R. Zhu, P. Hitzler, C. Shimizu, K. Currier, M. Schildhauer, et al., The S2 hierarchical discrete global grid as a nexus for data representation, integration, and querying across geospatial knowledge graphs, preprint arXiv:2410.14808 (2024).

[21] MSCI Inc., The global industry classification standard, https://www.msci.com/indexes/index-resources/gics, 2025.

[22] P. Hitzler, B. Parsia, P. Patel-Schneider, S. Rudolph, OWL 2 Web Ontology Language Primer (Second Edition), https://www.w3.org/TR/owl2-primer/, 2012. Https://www.w3.org/TR/owl2-primer/.

[23] Protege, Protégé, https://protege.stanford.edu/, 2025.

[24] Ontotext, Ontotext GraphDB, https://graphdb.ontotext.com/, 2025.