Proceedings of the ASME 2024
International Design Engineering Technical Conferences and
Computers and Information in Engineering Conference
IDETC/CIE2024
August 25-28, 2024, Washington, DC, USA

IDETC2024-143710

A BASIC FORMAL ONTOLOGY-BASED ONTOLOGICAL MODELING FOR PLAN AND OCCURRENCE, A BIOMANUFACTURING PROCESS VERIFICATION USE CASE

Dušan Šormaz[1], Saruda Seeharit[1], Boonserm Kulvatunyou[2], Miloš Drobnjaković[3]

[1] Department of Industrial and Systems Engineering
Ohio University, Athens, OH, USA
[2]Engineering Laboratory, National Institute of Standards and Technology
Gaithersburg, MD, USA
[3]Associate, Engineering Laboratory, National Institute of Standards and Technology
Gaithersburg, MD, USA

ABSTRACT

Biomanufacturing has gained significant importance in recent years due to its role in developing new medications, handling pandemics, and increasing the well-being of human populations. The nature of biochemical processes requires complex planning and control, with many controlled and non-controllable variables that impact the quality of bioproducts. Representing biomanufacturing process knowledge, control models, and actual occurrences in coherent ontologies could aid both humans and computers in dealing with the complexity. However, there is a lack of such coherent ontologies. Even though the Industrial Ontology Foundry (IOF) Core ontology has provided a groundwork based on the widely used Basic Formal Ontology (BFO) for such ontological requirements, there are still insufficient constructs and clear guidance on the representation of digital artifacts and their correspondences to the physical counterparts. This paper presents a framework to extend the IOF Core to address the gap. The framework is founded on establishing a counterpart (CR) relation pattern presented in our previous paper. Counterpart relation was selected for its ability to facilitate a more intuitive and concise representation of many kinds of digital artifacts (e.g., planned, designed) and physical entities (e.g., planning process, manufacturing process). We validated the approach with a process verification of a fed-batch bioreactor operation. The paper started by defining the use case requirement, which was followed by an ontology development. A knowledge graph of the bioprocess plan and occurrences of processes in the plan was then instantiated. Competency questions were used to concretize the ontology requirement from the use case, and subsequently, an executable set of queries was created from them and was used to computationally validate the ontology against the requirement. The GraphDB tool was used to support the validation. The result of this research not only showed that the CR pattern described in our previous paper could satisfy the requirements related to the digital thread of digital and physical process information, but it also demonstrated that several visualization approaches on graph data can be used to address competency questions. These findings provide insights into the future of data integration and management within biomanufacturing, highlighting the role of ontologies for improved data interoperability and analysis.

Keywords: Data/Information modeling, Design integration, Engineering informatics, Intelligent manufacturing, Manufacturing planning, Biomanufacturing

1. INTRODUCTION

Modern manufacturing has been experiencing technological advancements on both the hardware and software fronts. On the hardware side, real-time sensing through the Internet of Things (IoT) poises to allow for more timely decision-making and actions. The development of new manufacturing methods, such as additive manufacturing and continuous biomanufacturing, poses new challenges for understanding the process not seen in traditional discrete and batch manufacturing. These are challenges that require more connected data and domain knowledge residing in various sources including automation, sensing, and enterprise software tools (such as CAD, CAM, Simulation, PLM, MES, and MOM) which are used to design and plan products and processes, and to execute, monitor, and verify manufacturing operations and qualities. The need for a seamless connection between data and knowledge faces the problem of interoperability among these software and hardware tools. Ontologies have been proposed as a method to organize knowledge and provide a common language for communication between different tools, as they disambiguate semantic relations between various physical and digital entities, especially when bridging across operational areas and industry sectors [1,2].

The formalization of various types of digital entities and their relations to the various types and phases of physical entities has become increasingly important to accommodate the ultimate goal of building digital twins of products, processes, and systems [3]. Along these lines, the Industrial Ontology Foundry (IOF) has been founded to provide reference ontologies for various domains of manufacturing. The IOF community has recently released its Core module, which is based on the ISO 21838-2 Basic Formal Ontology (BFO) standard, as a foundation for other manufacturing domain ontologies [4]. The current work on the IOF Core did not include comprehensive models for digital, especially future artifacts. This is a challenge because BFO is a "realist ontology" [5]. However, there is a need to represent the details of future artifacts that are results of the work of designers and engineers (for example designs, process plans) with full semantic relationships to actual entities. In our previous paper, we have evaluated various approaches that extend IOF Core and BFO to address this need [6]. We demonstrated that the counterpart relation (CR) approach provides the best option using a discrete manufacturing use case scoped to capturing the correspondence between the design and the respective as-manufactured product. In this paper, we evaluate the same pattern as it applies to biomanufacturing process performance evaluation in terms of quality and compliance. Biomanufacturing is an important use case as connected and highly accessible data and knowledge are becoming an important tool in advancing it [7]. Biomanufacturing is currently undergoing a transition towards Industry 4.0 including new production modalities and employment of advanced control methods, all requiring hybrid, data-driven support. Consequently, there is an emerging need for greater interoperability and data contextualization across different process lifecycle stages.

The rest of the paper is organized as follows. Section 2 presents previous works in the area of ontologies and their applications to manufacturing. Section 3 describes the methodology that was applied in the research. Section 4 explains the biomanufacturing use case and its competency questions, while Section 5 shows the validation of the approach on the use case. The paper closes with Section 6 which presents the conclusion and future work.

2. PREVIOUS WORK

An ontology encodes domain's concepts and relationships with a standardized vocabulary and axiomatic expressions, enabling computer understanding of domain knowledge. OWL, recommended by W3C for the Semantic Web, is the preferred ontology encoding language, grounded in Description Logics (DL) [8]. OWL extends RDF (Resource Description Framework) by offering a more expressive axiomatic representation of knowledge, enabling advanced modeling and reasoning across data and domain knowledge. A DL reasoner infers logical outcomes from stated facts and axioms. Using it with graph data querying engine they assist in information classification, data inconsistency debugging, and knowledge discovery.

Ontologies have been instrumental in unifying biology through the standardization of concepts and terminologies across various subfields, as evidenced by the Gene Ontology's success in providing a controlled vocabulary for gene products across multiple organisms [9]. In addition, the application of ontologies has facilitated the specification and constraint of values within biological databases enabling more advanced data integration and analysis [10].

Because of the successes of Gene Ontology in the biological research domain, several literatures pointed out the need for a coherent suite of ontologies to advance biomanufacturing. Generally, ontology models allow the representation of knowledge in a structured form that can assist humans in development, optimization, and compliance of biomanufacturing processes. Drobnjaković et al. specifically highlighted the needs and benefits of applying ontology to the biomanufacturing sector. They indicated that it could help advance the sector's goal of efficient, intelligent, and sustainable production systems by aiding in the developments of hybrid-data-driven digital twins [6,7,11]. Chen et al. [12] concurred, stating that the emerging significance of accurate digital representations, an ontology's forte, is necessary for optimizing manufacturing processes within pharmaceutical and biopharmaceutical manufacturing. Complementarily, Smith and Ceusters [13] provided foundational insights into the Information Artifact Ontology, focusing on the conceptualization of digital artifacts, which underpins the effective implementation of digital twins. They emphasized the critical aspect of 'aboutness' in digital representations, which is generally known as digital thread.

Building on these insights, the Industrial Ontologies Foundry (IOF) seeks to replicate the success seen in other sectors by fostering the development of interoperable ontologies for the entire industrial manufacturing [2]. By using the Basic Formal Ontology (BFO) as a basis, the IOF's efforts are geared towards establishing a unified ontology framework that enhances data interoperability and supports the manufacturing domain's evolving needs [14]. Since the release of the IOF Core Beta in 2022 and subsequently full release in 2023, several papers have shown its broad applicability in providing the basis for interoperability and analytical usages [15–19]. Drobnjaković et al. [11] and Nikolov et at. [20], in particular, showed IOF Core used in the biomanufacturing domain. The former showed how the IOF Core can be extended to represent data and information flow between a digital twin and a fed-batch biomanufacturing process. The latter showed the extension of IOF Core to model the life cycle assessment terminology from ISO 14040 and how it can be used to connect various sources of data for CO2 emissions analysis of the fed-batch bioreactor manufacturing process. Advancing from our foundational research which compared ontological approaches for digital artifact representation and their connections to physical objects, this study focuses on further evaluation of the Counterpart Relation (CR) approach, which showed advantages over others in [6]. Originally proposed in [3], it provides a structured framework that includes patterns and counterpart relations that distinguish digital from physical entities while allowing for correspondences. For example, when a jet engine has an *Counterpart relation* to a another jet engine, this indicates that the former is a digital entity while latter is a physical entity. In other words, detail of the former entity like dimension is imaginary, expected, or pattern, while detail of the latter is real, measured, and physical. However, modern manufacturing enterprises use multitudes of digital representations to represent an entity (object or process) throughout its lifecycle stages from requirement, design, analyses, to the eventual disposal or recycling. Our exploration into the CR approach aims to further validate its scalability across these lifecycle stages.

3. METHODOLOGY

This section outlines the methodology employed in applying and validating the Counterpart Relation (CR) ontology approach for representing plan and process. First, the CR approach is described in more detail including how it is realized in the IOF ontology. Subsequently, the approach to validate it using a biomanufacturing use case is described.

3.1 Realization of the CR approach in the IOF ontology

The CR methodology was grounded on the comprehension that a product's lifecycle encompasses several stages. By aligning the CR approach with the IOF, it is ensured that CR not only fits within the IOF framework but also supports the goal of enhancing data interoperability across product life cycle stages.

The CR approach can be applied on planned future processes and actual processes using relations shown in **FIGURE 1**. The figure shows that the planning process *pp1* has a plan specification *ps1* as its output and is related to a future planned process *cgp1-p* using *plansProcess* relation. The plan specification *ps1 prescribesPlannedEntity cgp1-p*, and also *prescribesActualEntity cgp1* (which is an instance of actual planned process that is happening or that already happened). The relation *isCounterpartOf* is established between *cgp1* and *cgp1-p*, to signify that the actual process *cgp1* should have the same participants, duration, and process characteristics as *cgp1-p* which are all prescribed during the planning process *pp1* and recorded in the plan specification *ps1*.

The above-mentioned relations have been implemented in a CR ontology file and integrated within IOF framework for the purpose of this study.

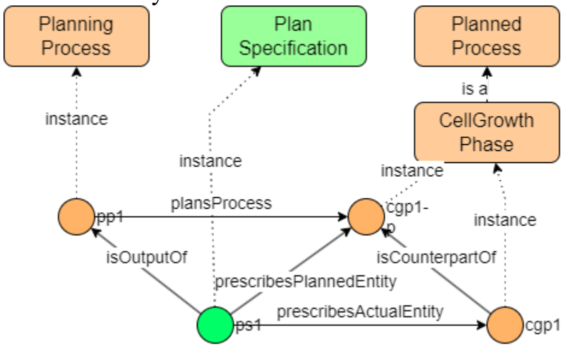


FIGURE 1: RELATIONS BETWEEN PLANNED FUTURE PROCESS AND ACTUAL PROCESS[1]

3.2 Ontology Development and Validation

The methodology for ontology development and validation consists of the following steps: a) use case description, b) competency question elicitation, c) ontology class and property extension to the IOF core ontology, d) generation of knowledge graph with instance data, e) SPARQL queries encoding to reflect the competency questions, and f) execution and verification of queries and results. A realization of each of these steps is demonstrated in each subsection of section 4.

4. CASE STUDY

The process understanding plays a critical role in advancing the biomanufacturing of bioindustrial materials and biopharmaceutical products. Characterized by their complex processes and strict control requirements, knowledge-based combined with data-driven approaches can assist humans and computer to better understand and control the unit processes.

In our collaboration with industry partners, process monitoring and verification of fed-batch bioreactor unit operation in batch manufacturing shown in **FIGURE 2** is a critical capability. It exemplifies the technical challenges and precision needed for successful biopharmaceutical production. Therefore, we use the unit operation as a use case to validate the CR approach in this paper.

4.1 Use Case Description

The fed-batch production bioreactor unit operation (buo) is structured into two critical phases: the growth phase (gp) and the production phase (pp), sequentially. It is imperative that the pH level is maintained at 7 ± 0.1 throughout both phases. To ensure precise pH control, a pH controller with a precision capability of 0.1 is employed as an integral process participant. The operational protocol specifies the choice between *bioreactor1* and *bioreactor2*.

---

[1] Orange-colored shapes are subclasses or instances of BFO Process, while green-colored shapes are subclasses or instances of IOF Information Content Entity (ICE), which is, in turn, a subclass of BFO Generically Dependent Continuant (GDC). It is important to note that GDC cannot bear any quality (e.g., color, length); therefore, capturing all the quality specifications with ICE would necessitate the creation of a subclass hierarchy under ICE that is essentially the same as one needed under the BFO Quality. This is explained in more detail in [6].

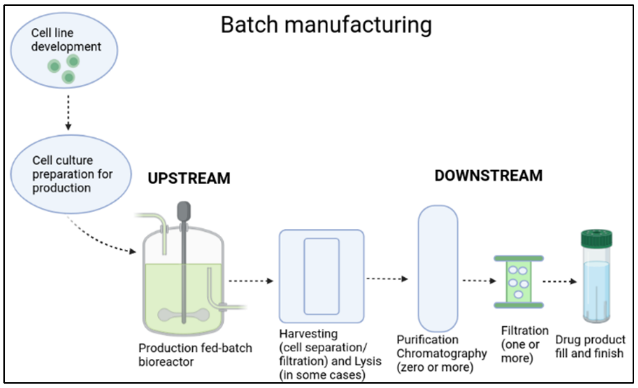


FIGURE 2: BATCH BIOMANUFACTURING (adapted from [7])

Both bioreactors have the same capacity of 3 liters. The intended duration for the unit operation spans 21 days. The growth phase is initially scheduled to last 5 days, followed by the production phase, which is planned to last 16 days. While the real process includes continuous monitoring of pH level (another type of process) that could introduce additional ontological challenge, we left that out of scope in this initial phase of research and focused on the dynamic between detailed planning and its physical execution. Despite rigorous planning to maintain optimal conditions through two critical phases, unforeseen complications may lead to an early termination, short of the planned 21 days. This discrepancy underscores the inherent unpredictability of biological systems and highlights the importance of adaptability in biomanufacturing.

For the purpose of validation, the use case has 5 hypothetical production runs, of which the first four are on bioreactor 1 and the last one is on bioreactor 2. The details of these production runs will be explained later in section 4.4.

4.2 Competency Questions

Competency Questions (CQs) serve as a pivotal tool in ontology development, functioning as structured inquiries that outline the scope and validate the effectiveness of an ontology in capturing domain-specific knowledge. They act as benchmarks to assess whether an ontology meets the desired requirements and can answer specific questions pertinent to its domain [21]. This concept, integral to ontology engineering, ensures that the developed ontologies are both relevant and functional in their respective contexts.

The key validation aspect is to ensure that correspondences between various digital and actual entities and their characteristics (such as participants, time, dimensions, or other physical properties) that may not be directly associated with the entities can be established, traced, and compared. The key design constraints are consistent design patterns for establishing those correspondences, consistent with the IOF and BFO view of the universe of discourse, minimizing the modeling burden including the number of classes and properties,

Based on the detailed description of our process, we developed a series of competency questions (CQs) to guide our ontology development and validation and enhance our understanding of the operational dynamics:

A. Which equipment or pieces of equipment were utilized in an actual process?
B. Which equipment or pieces of equipment were allowed in the operation plan?
C. What is the difference of duration between an actual process and its plan?
D. What are the minimum and maximum pH values recorded during the actual process?
E. Did the pH remain within the specified limits throughout the entire process?
F. What precision setting was the pH controller configured to?

G. Which processes unfolded as planned?
H. Was the production phase initiated in the specified run?
I. Did the production phase commence on the scheduled day?
J. Is the implemented pH controller's precision in alignment with the process requirements?
K. Which phases proceeded according to the initial plan?

These competency questions pose several requirements that the model needs to satisfy:

- There is a need to establish the correspondences between a process plan, scheduled occurrence, and actual occurrence,
- The process plan needs to capture data about the process participants,
- The process plan needs to capture data about subprocesses and their sequence,
- The process plan needs to capture data about process durations, and any overlapping needs,
- The process plan needs to capture data about inputs and outputs of each phase (subprocess).

Those requirements signify the need to extend IOF Core classes to represent various types of processes, its subprocesses, qualities, information content, and participants that are best captured by correspondence between the ideal planned process and actual planned process. These will be shown in the following sections.

4.3 IOF Core Extension Development

The ontology development encompasses importing the IOF and CR ontologies, augmenting them with the necessary classes, properties, axioms, and individuals identified by the use case description and competency questions. **FIGURE 3** illustrates the extension – only additional classes are needed. The graph includes both existing and newly added classes pertinent to the use case. For enhanced clarity, existing BFO classes are shown as magenta rectangles, existing IOF classes in cyan, and the added domain, use-case-specific classes in orange, as detailed in the legend. Existing classes are in the top part of the figure. Added classes extending the existing ones are naturally at the bottom part. An ontology according to this figure is encoded in OWL and saved into an RDF file using Protégé.

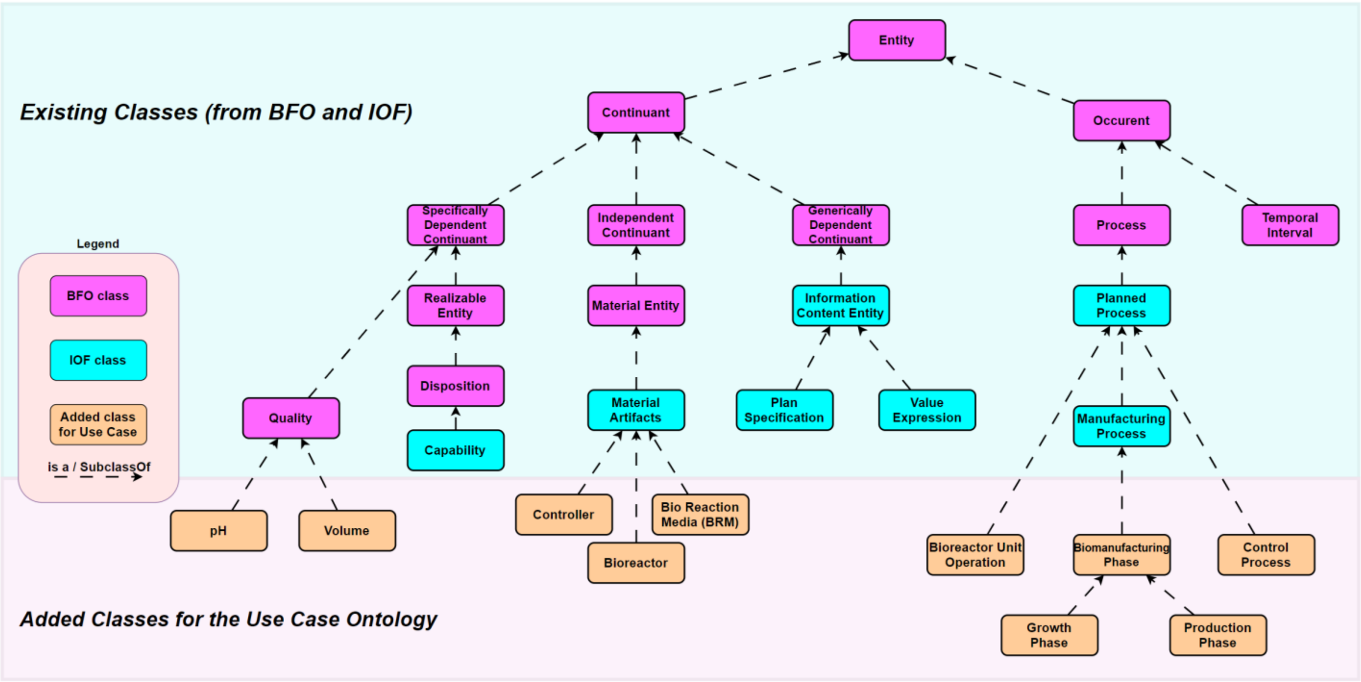


FIGURE 3: DEPICTION OF THE USE CASE-SPECIFIC CLASSES, EXPANDING UPON THE FOUNDATIONS OF IOF AND BFO

4.4 Encoding the Knowledge Graph

From the use case description in section 4.1 and domain ontology given in **FIGURE 3**. We have created instances (individuals) for planned and actual entities as follows.

For planning entities, there is one plan specification *ps1*, which specifies two alternate processes, *buo1-d* and *buo2-d* (see **FIGURE 4**). The primary difference between these processes is that *buo1-d* is designed to use bioreactor *b1*, while *buo2-d* is specified to use bioreactor *b2*. Both of those processes have their subprocesses related to the grow phase, *cgp1-d* and *gp2-d*, and production phase, *pp1-d* and *pp2-d*. The details about expected pH values, and controller are also shown in the figure. The naming convention for the individuals in the figure is as follows. The ideal situation representing the plan has suffix *-d* at the end of their names, e.g., *buo1-d* is an individual 1 of a 'Bioreactor Unit Operation' plan and *pH-d* is its ideal pH value range. The executed process does not have a suffix, e.g., *buo2* is an individual for an executed 'Bioreactor Unit Operation' and *pH1* is a pH measurement during the process run.

Among the actual runs, we have generated five hypothetical production runs and developed the following scenarios for testing the ontology:

1. **Production Run 1:** The production phase ended earlier than anticipated, at day 18, which will be individual *buo1;*
2. **Production Run 2:** The growth phase extended by 2 days (totaling 7 days), consequently shortening the production phase to 14 days, *buo2*;
3. **Production Run 3:** The pH went out of range and hence a failure occurred during the growth phase, preventing the start of the production phase, *buo3*;
4. **Production Run 4:** The process unfolded exactly as planned, *buo4*;
5. **Production Run 5:** The process was terminated on day 10, *buo5*.

It should be noted that processes *buo1* – *buo4* use bioreactor *b1*, whereas *buo5* utilizes bioreactor *b2*. Consequently, only *buo5* has *buo2-d* as its plan, highlighting its unique requirements and setup. The remaining runs, from *buo1* to *buo4*, adhere to the *buo1-d* plan. This also implies that *buo5* can be executed in parallel with other processes, but the reasoning about the sequences is beyond the scope of the paper.

A portion of the resulting knowledge graph covering **buo1** run is shown in **FIGURE 4**. The reader should notice the application of the Counterpart Relation (CR) approach through the *hasCounterpart* relation between the ideal planned processes and the actual/executed ones. All individuals for the two ideal planned processes, *buo1-d* and *buo2-d* are shown (as circles), including their subprocesses (in dark beige), the participants (in pink), qualities (in blue), and plan specifications (in green). One actual process, *buo1* is shown with its subprocesses and all three of them are connected via *hasCounterpart* to their respective plans, *buo1-d*, and its subprocesses, *gp1-d* and *pp1-d*, with counterpart relations. The plan, *buo2-d*, is not connected with any actual processes, because it does not prescribe *buo1*. The knowledge graph about the plan specification, planned processes, and the five hypothetical production runs **buo1** – **buo5** has been encoded using Protégé according to the ontology created in section 4.3 that imports the CR ontology. **FIGURE 5**. includes all planned process individuals representing both the plans and the actual production runs.

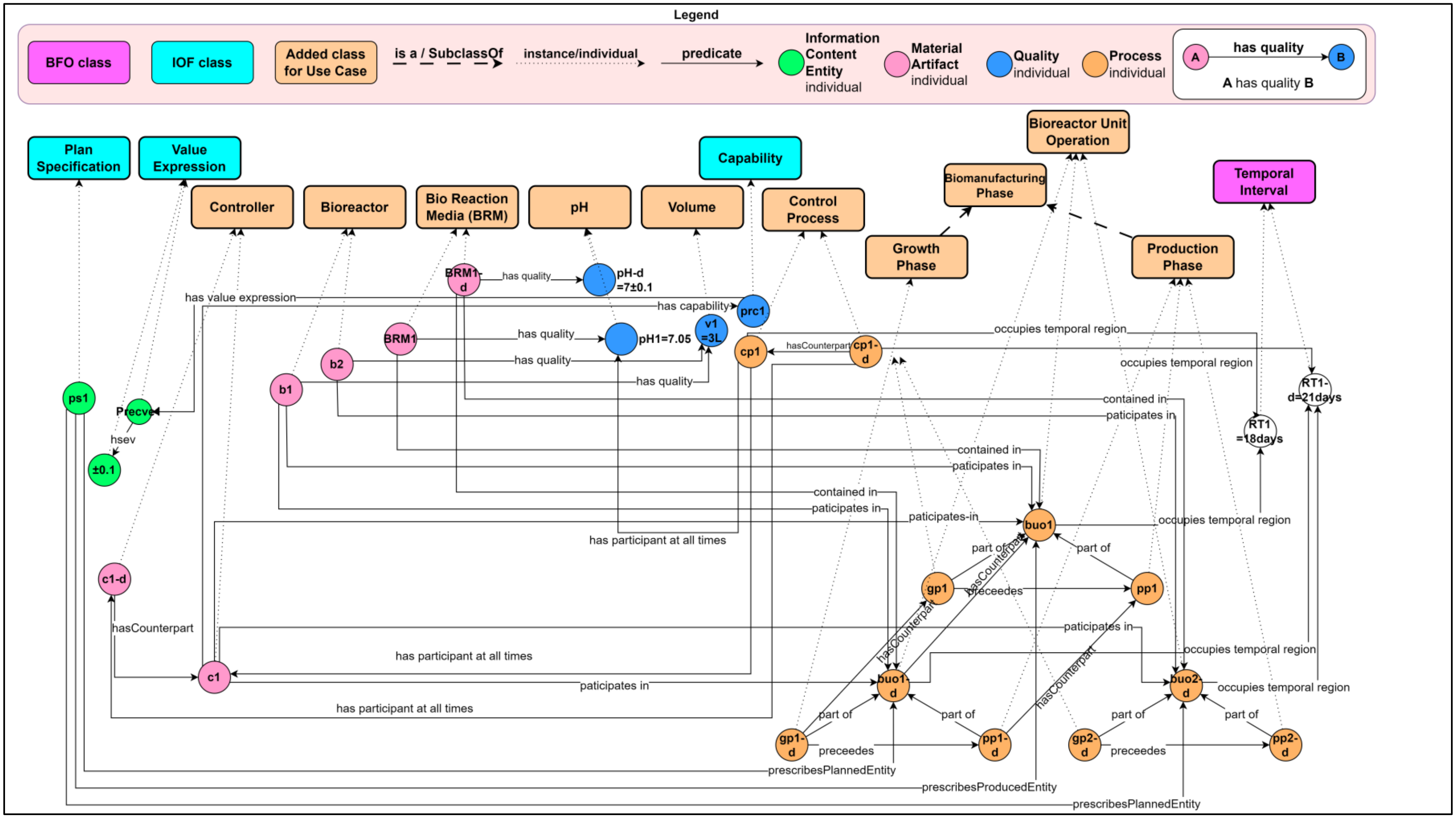


FIGURE 4: APPLYING CR APPROACH TO THE FED-BATCH PRODUCTION BIOREACTOR OPERATIONS.

In the next subsection, we show the formulation of competency questions using SPARQL queries according to the ontology.

4.5 SPARQL Query Formulation for CQs

To validate these CQs, we leveraged SPARQL queries as a tool for interrogating the underlying knowledge graph, ensuring our ontology's constructs can respond according to the required semantics and relationships.

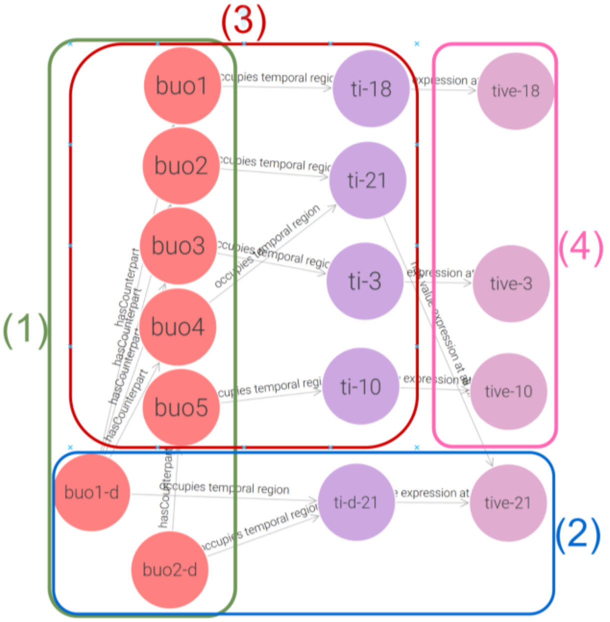


FIGURE 5: KNOWLEDGE GRAPH REPRESENTING THE EXECUTION ORDER OF SPARQL QUERY FOR QUESTION C

In this paper, we illustrate the use of SPARQL queries on only questions C and K as they are more insightful for the validation. The competency question K, "Which phases proceeded according to the plan?", encompasses a broad inquiry into the process's adherence to the plan, including not just the sequence and execution of various phases but also specific operational parameters like duration and pH levels. Within this broader context, question C, "What is the difference of duration between actual and plan?" serves as a critical sub-question, focusing on the aspect of time taken and its alignment with the planned schedule.

To comprehensively answer question K, it is essential to evaluate the duration of each phase as outlined in question C, then verify the sequence of each phase and ensure that all planned equipment was used throughout the process.
This approach should allow for a detailed validation of the process's execution against its planned blueprint, incorporating both temporal and qualitative metrics.

Key sections of a SPARQL query include:

- **PREFIX:** This section defines the abbreviations for the URIs used in the query, simplifying the query structure, and improving readability.
- **SELECT:** Specifies the variables that will be returned by the query. It can include conditions to filter the results.
- **WHERE:** Contains the pattern matching against the knowledge graph data. This is where the conditions for the data retrieval are defined.
- **FILTER:** Applies additional conditions to the query, refining the results based on specific criteria.
- **ORDER BY:** Determines the order in which the results are returned, based on one or more variables.

SPARQL query for competency question C is shown in **FIGURE 6**. A pivotal aspect of our SPARQL query design involves the predicate *cr:hasCounterpart,* which significantly streamlines the query process. By focusing directly on the relationship between the plan and the executed processes (see bolded portion in the figure), this predicate eliminates the cumbersome need to incorporate extensive plan specifications and multiple prescribe relations (see [6] for details).

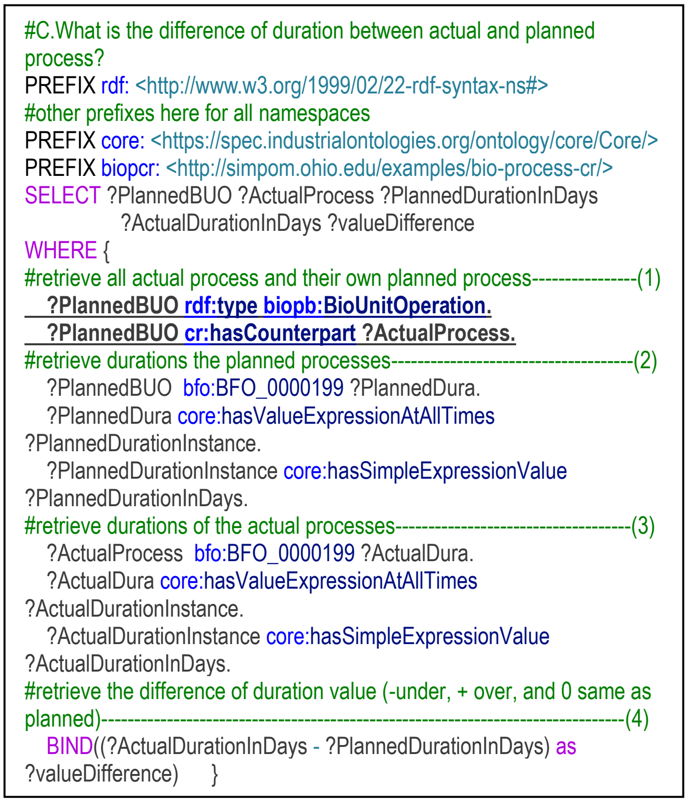

```
#C.What is the difference of duration between actual and planned
process?
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
#other prefixes here for all namespaces
PREFIX core: <https://spec.industrialontologies.org/ontology/core/Core/>
PREFIX biopcr: <http://simpom.ohio.edu/examples/bio-process-cr/>
SELECT ?PlannedBUO ?ActualProcess ?PlannedDurationInDays
         ?ActualDurationInDays ?valueDifference
WHERE {
#retrieve all actual process and their own planned process----------------(1)
  ?PlannedBUO rdf:type biopb:BioUnitOperation.
  ?PlannedBUO cr:hasCounterpart ?ActualProcess.
#retrieve durations the planned processes-------------------------------------(2)
  ?PlannedBUO  bfo:BFO_0000199 ?PlannedDura.
  ?PlannedDura core:hasValueExpressionAtAllTimes
?PlannedDurationInstance.
  ?PlannedDurationInstance core:hasSimpleExpressionValue
?PlannedDurationInDays.
#retrieve durations of the actual processes-----------------------------------(3)
  ?ActualProcess  bfo:BFO_0000199 ?ActualDura.
  ?ActualDura core:hasValueExpressionAtAllTimes
?ActualDurationInstance.
  ?ActualDurationInstance core:hasSimpleExpressionValue
?ActualDurationInDays.
#retrieve the difference of duration value (-under, + over, and 0 same as
planned)-----------------------------------------------------------------------------(4)
  BIND((?ActualDurationInDays - ?PlannedDurationInDays) as
?valueDifference)     }
```

FIGURE 6: SPARQL QUERY FOR COMPETENCY QUESTION C

This strategic simplification not only reduces the query's complexity but also enhances the clarity and directness of our ontology validation efforts. Through the application of these SPARQL queries, we not only validate the Counterpart Relation (CR) approach's effectiveness in linking the intricate relationships between digital and physical entities but also underscore its potential in streamlining ontology-based data querying processes. This is visible in **FIGURE 5**, which shows the portion of knowledge graph utilized in the query.

The numbers 1, 2, 3, and 4 in the figure correspond to the portions of the query labeled as 1 – matching of the processes representing the plans and their actuals, 2 – determining the duration of processes in the plan, 3 – finding the durations of actual processes, and 4 – getting their values from data properties.

This methodological approach, underpinned by precise and structured query execution, illuminates the practical strengths of our ontology in addressing complex real-world challenges. Given that the competency question K, "Which phases proceeded according to the initial plan?", its query as shown in **FIGURE 7** has a filter that sets all three for those conditions to TRUE. These conditions are: 1) *hasCorrectPhaseSequence* (ensuring the growth phase precedes the production phase), 2) *hasCorrectPhaseDuration* (where the durations match the planned durations), and 3) *hasRequiredEquipment* (accounting for all necessary equipment specified by a particular predicate). The light (yellow) rectangles emphasize the relationships developed in the CR ontology for those purposes.

5. Validation of the Results

This section presents results of the validation by executing the queries on the knowledge graph. For executing these SPARQL queries, we employed GraphDB[2], an RDF compliant knowledge graph tool [22]. Before running the queries, we first imported all the related ontology files from Protégé into the server. While we executed all queries for all competency questions mentioned in section 4.2 and the results were as expected (i.e., the ontology satisfies the use case requirements), this section displays results only for the competency questions C and K. By default, the query results are shown in a tabular format, but they can be presented in various formats for better visualization of the process compliance.

```
SELECT ?ActualPhases ?hasCorrectPhaseSequence ?hasCorrectPhaseDuration ?hasRequiredEquipment
((?hasCorrectPhaseSequence && ?hasCorrectPhaseDuration && ?hasRequiredEquipment ) as ?OccurredAccordingToplan)
WHERE {
    {SELECT ?ActualPhases
                (count(distinct(?PlannedFollowingSteps)) as ?numberOfPlannedFollowingSteps)
                (count(distinct(?RealFollowingSteps)) as ?numberOfRealFollowingSteps)
                ?hasCorrectPhaseSequence
    WHERE{   ?has_participant_all rdfs:label "has participant at all times"@en.
            ?prop_occurr_part rdfs:label "has proper occurrent part"@en.
            ?ProcessPlan cr:prescribesPlannedEntity ?PlannedProcess.
            ?ProcessPlan cr:prescribesActualEntity ?ActualProcess.
                ?PlannedProcess ?prop_occurr_part ?PlannedPhases.
                ?PlannedPhases cr:hasCounterpart ?ActualPhases.
            ?ActualPhases rdf:type/rdfs:subClassOf biopb:BioManufacturingProcess.
#check phase sequence
            OPTIONAL{?PlannedPhases bfo:BFO_0000063+ ?PlannedFollowingSteps.}
            OPTIONAL{?ActualPhases bfo:BFO_0000063+ ?RealFollowingSteps.}
        } GROUP BY ?ActualPhases ?hasCorrectPhaseSequence}
        BIND (?numberOfPlannedFollowingSteps=?numberOfRealFollowingSteps as ?hasCorrectPhaseSequence).
# Fetch the required data.
    ?has_participant_all rdfs:label "has participant at all times"@en.
    ?prop_occurr_part rdfs:label "has proper occurrent part"@en.
    ?ProcessPlan cr:prescribesPlannedEntity ?PlannedProcess.
    ?PlannedProcess ?prop_occurr_part ?PlannedPhases.
    ?PlannedPhases cr:hasCounterpart ?ActualPhases.
    ?ActualPhases rdf:type/rdfs:subClassOf biopb:BioManufacturingProcess.
# Equipment section.
    ?PlannedPhases ?has_participant_all ?RequiredEquipment.
    ?ActualEquipment rdf:type/rdfs:subClassOf core:MaterialArtifact.
# Check equipment
    ?ActualPhases ?has_participant_all ?ActualEquipment.
    BIND ((?RequiredEquipment = ?ActualEquipment)
    || (EXISTS {?ActualEquipment cr:isCounterpartOf ?RequiredEquipment}) AS ?hasRequiredEquipment)
# Duration Section.
    OPTIONAL {
    ?PlannedPhases bfo:BFO_0000199/core:hasValueExpressionAtAllTimes/core:hasSimpleExpressionValue ?PlannedPhaseDurationInDays.
    ?ActualPhases bfo:BFO_0000199/core:hasValueExpressionAtAllTimes/core:hasSimpleExpressionValue ?ActualPhaseDurationInDays.
    BIND ((?ActualPhaseDurationInDays = ?PlannedPhaseDurationInDays) as ?hasCorrectPhaseDuration).
    }
# Show only process that occurred according to plan --> every condition is true
FILTER(?hasCorrectPhaseSequence= true && ?hasCorrectPhaseDuration= true && ?hasRequiredEquipment= true)
}
GROUP BY ?ActualPhases ?hasCorrectPhaseSequence ?hasCorrectPhaseDuration ?hasRequiredEquipment
```

FIGURE 7: SPARQL QUERY FOR COMPETENCY QUESTION K

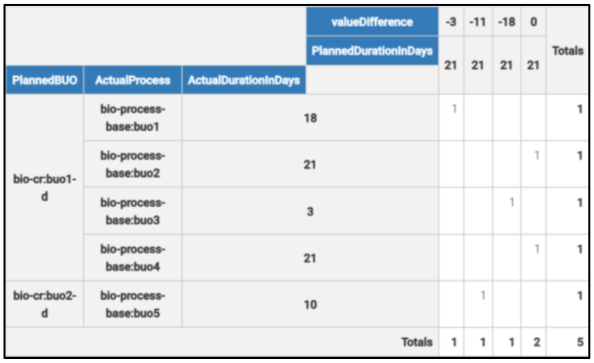
---

[2] https://www.ontotext.com/products/graphdb

5.1 Query C “What is the difference of duration between actual and plan?”

Based on the question, we found that pivot table is an effective visualization for conveying both compliance and conflict between the plans and the actual processes. The outcome of the query is displayed in a pivot table shown in **FIGURE 8**. Specifically, the *totals* row, positioned at the bottom, reveals that there are two actual processes that have correct process durations by comparing the execution process to its planned counterpart durations.

For this query C on executed result, we selected a vertical bar chart to illustrate the difference of durations in days between the executed process and its planned counterpart as shown in **FIGURE 9**. The presence of a longer red bar pointing downwards effectively underscores the number of days by which the executed process was shortened compared to its original plan. This visual cue serves to immediately draw attention to the discrepancy between planned and actual timelines, emphasizing the deviation in a clear and impactful manner.



		valueDifference	-3	-11	-18	0	Totals
		PlannedDurationInDays	21	21	21	21	
PlannedBUO	ActualProcess	ActualDurationInDays					
bio-cr:buo1-d	bio-process-base:buo1	18	1				1
	bio-process-base:buo2	21				1	1
	bio-process-base:buo3	3			1		1
	bio-process-base:buo4	21				1	1
bio-cr:buo2-d	bio-process-base:buo5	10		1			1
		Totals	1	1	1	2	5

FIGURE 8: RESULTS OF THE QUERY C IN A PIVOT TABLE

5.2 Query K “Which phases proceeded according to the initial plan?”

The filter, highlighted by a blue rectangle at the bottom of **FIGURE 7**, was intentionally omitted to observe the subsequent results and compare different scenarios. The results of this adjusted query (K) are then presented in **FIGURE 10**, where the pivot table adopts a "Count as Fraction of Total" approach. The resulting percentages provide insightful observations. The table's columns represent the true or false status of each condition. The analysis reveals that all the planned processes have their required equipment in place. Furthermore, approximately 11.1% of all phases (1 out of 9) display incorrect phase sequences and fail to meet the correct phase duration. Additionally, around 44.4% (4 out of 9) of all phases have the correct phase sequences but incorrect phase durations. Lastly, 44.4% (4 out of 9) of all phases are found to have proceeded according to their initial plans, indicating a significant portion of the project adheres to its original timeline and sequence criteria. The Counterpart Relation (CR) approach helps facilitate the systematic formalization of the queries in order to validate the competency questions and uncover hidden knowledge.

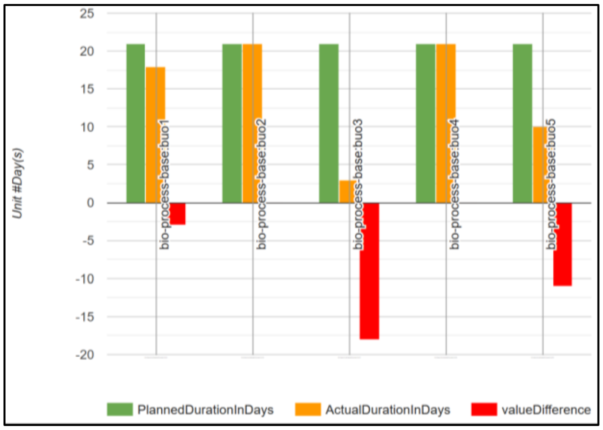
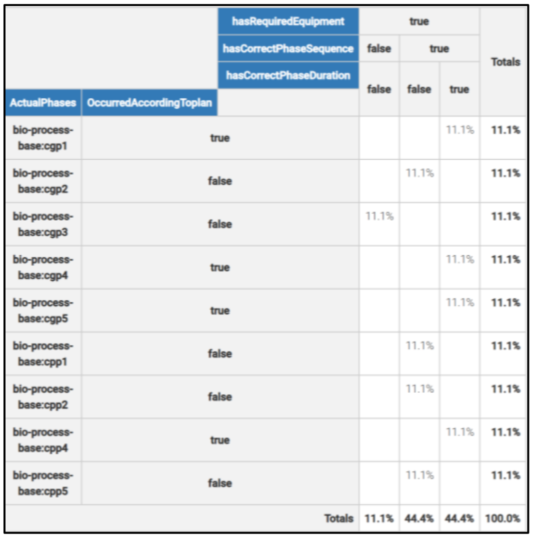


FIGURE 9: BAR CHART DEPICTING THE DIFFERENCES BETWEEN PLANNED AND ACTUAL PROCESS DURATIONS



	hasRequiredEquipment	true			Totals
	hasCorrectPhaseSequence	false	true		
	hasCorrectPhaseDuration	false	false	true	
ActualPhases	OccurredAccordingToplan				
bio-process-base:cgp1	true			11.1%	11.1%
bio-process-base:cgp2	false		11.1%		11.1%
bio-process-base:cgp3	false	11.1%			11.1%
bio-process-base:cgp4	true			11.1%	11.1%
bio-process-base:cgp5	true			11.1%	11.1%
bio-process-base:cpp1	false		11.1%		11.1%
bio-process-base:cpp2	false		11.1%		11.1%
bio-process-base:cpp4	true			11.1%	11.1%
bio-process-base:cpp5	false		11.1%		11.1%
	Totals	11.1%	44.4%	44.4%	100.0%

FIGURE 10: QUERY K - RESULT IN A PIVOT TABLE

6. CONCLUSIONS

The paper has presented an ontological approach for biomanufacturing process verification based on establishing the Counterpart Relation between processes representing a plan and its actuals. The resulting ontology model provides unambiguous reasoning about the actual process compliance to its plan in three categories: durations, participants, and subprocess sequences. Simultaneously, the approach provides uniform semantic relations for planned and actual properties. The approach is demonstrated on the simplified, 2-phase biomanufacturing unit operation use case that requires compliance checking of participants, sequences, and durations. The competency questions from the use case were transformed into SPARQL queries, which were then effectively run on knowledge graph software, thereby validating the effectiveness of our approach.
Results presented here indicate that the integration of ontological models, particularly through the lens of the CR approach, offers a method for capturing the lifecycle of biomanufacturing processes. This method significantly eases the process of systematically delineating the connections between digital

planning and physical execution that can be extended to other process analysis.

Although the results are encouraging, further research is reasonable to assess the approach's applicability and scalability. Future studies should explore problems where combinations of multiple digital artifacts are present and provide tracking of the phases of product lifecycle (such as requirements, design, planning, simulation, scheduling, and actual production) with measurements of performance in each phase.

DISCLAIMER AND ACKNOWLEDGEMENTS

The research presented in this document was supported in part by funding from the National Institute of Standards and Technology (NIST) through a cooperative agreement with Ohio University, identified by the agreement number 70NANB22H155. It is important to note that this paper references certain commercial software products solely for the purpose of demonstration. This mention should not be interpreted as an endorsement or approval by NIST, nor should it be considered a declaration that these products are the best available options for their respective uses.

REFERENCES

[1] Smith, B., Ameri, F., Cheong, H., Kiritsis, D., Sormaz, D., Will, C., and Otte, J. N., 2019, "A First-Order Logic Formalization of the Industrial Ontology Foundry Signature Using Basic Formal Ontology."

[2] Kulvatunyou, B., Wallace, E., Kiritsis, D., Smith, B., and Will, C., 2018, "The Industrial Ontologies Foundry Proof-of-Concept Project," *Advances in Production Management Systems. Smart Manufacturing for Industry 4.0*, I. Moon, G.M. Lee, J. Park, D. Kiritsis, and G. Von Cieminski, eds., Springer International Publishing, Cham, pp. 402–409.

[3] Sormaz, D., and Sarkar, A., 2021, "Interoperability between PLM, ERP, and MES Systems Using Formal Ontologies," Taichung, Taiwan.

[4] Drobnjakovic, M., Kulvatunyou, B., Ameri, F., Will, C., Smith, B., and Jones, A., "The Industrial Ontologies Foundry (IOF) Core Ontology," Tarbes, FR.

[5] Merrell, E. C., Kelly, R. M., Kasmier, D., Smith, B., Brittain, M., Maki, E., Heisey, C. W., and Bush, K., "Benefits of Realist Ontologies to Systems Engineering."

[6] Šormaz, D., Kulvatunyou, B., Drobnjaković, M., and Seeharit, S., 2023, "Comparative Study of Approaches for an Ontology of Digital Artifacts," *Volume 2: 43rd Computers and Information in Engineering Conference (CIE)*, American Society of Mechanical Engineers, Boston, Massachusetts, USA, p. V002T02A057.

[7] Drobnjakovic, M., Hart, R., Kulvatunyou, B. (Serm), Ivezic, N., and Srinivasan, V., 2023, "Current Challenges and Recent Advances on the Path towards Continuous Biomanufacturing," Biotechnol. Prog., **39**(6), p. e3378.

[8] Abburu, S., "A Survey on Ontology Reasoners and Comparison," Int. J. Comput. Appl., **57**.

[9] Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G., 2000, "Gene Ontology: Tool for the Unification of Biology," Nat. Genet., **25**(1), pp. 25–29.

[10] Stevens, R., 2000, "Ontology-Based Knowledge Representation for Bioinformatics," Brief. Bioinform., **1**(4), pp. 398–414.

[11] Drobnjakovic, M., Kulvatunyou, B., Frechette, S., and Srinivasan, V., 2023, "Recent Developments in Ontology Standards and Their Applicability to Biomanufacturing," *Volume 2: 43rd Computers and Information in Engineering Conference (CIE)*, American Society of Mechanical Engineers, Boston, Massachusetts, USA, p. V002T02A058.

[12] Chen, Y., Yang, O., Sampat, C., Bhalode, P., Ramachandran, R., and Ierapetritou, M., 2020, "Digital Twins in Pharmaceutical and Biopharmaceutical Manufacturing: A Literature Review," Processes, **8**(9), p. 1088.

[13] Smith, B., and Ceusters, W., 2015, "Aboutness: Towards Foundations for the Information Artifact Ontology."

[14] Arp, R., Smith, B., and Spear, A. D., 2015, *Building Ontologies with Basic Formal Ontology*, MIT Press, Cambridge, MA, USA.

[15] Ameri, F., and Kulvatunyou, B., 2019, "Modeling a Supply Chain Reference Ontology Based on a Top-Level Ontology," *Volume 1: 39th Computers and Information in Engineering Conference*, American Society of Mechanical Engineers, Anaheim, California, USA, p. V001T02A052.

[16] Hodkiewicz, M., Low, E., and Woods, C., 2020, "Towards a Reference Ontology for Maintenance Work Management," CEUR Workshop Proc.

[17] Zheng, X., Lu, J., Arista, R., Hu, X., Lentes, J., Ubis, F., Sorvari, J., and Kiritsis, D., 2021, "Development of an Application Ontology for Knowledge Management to Support Aircraft Assembly System Design," *CEUR Workshop Proceedings*, Bolzano, Italy.

[18] Ameri, F., Sormaz, D., Psarommatis, F., and Kiritsis, D., 2022, "Industrial Ontologies for Interoperability in Agile and Resilient Manufacturing," Int. J. Prod. Res., **60**(2), pp. 420–441.

[19] Li, H., Wang, G., Lu, J., and Kiritsis, D., 2022, "Cognitive Twin Construction for System of Systems Operation Based on Semantic Integration and High-Level Architecture," Integr. Comput.-Aided Eng., **29**(3), pp. 277–295.

[20] Nikolov, A., Drobnjakovic, M., and Kulvatunyou, B., 2023, "Produce It Sustainably: Life Cycle Assessment of a Biomanufacturing Process Through the Ontology Lens," *Advances in Production Management Systems.*, Springer Nature Switzerland, Cham, pp. 504–517.

[21] Wiśniewski, D., Potoniec, J., Ławrynowicz, A., and Keet, C. M., 2019, "Analysis of Ontology Competency Questions and Their Formalizations in SPARQL-OWL," J. Web Semant., **59**, p. 100534.

[22] Ontotext, 2024, "GraphDB Documentation Release 10.6.2."