# An ontology for maintenance activities and its application to data quality

Caitlin Woods [a,*], Matt Selway [b], Tyler Bikaun [a], Markus Stumptner [b] and Melinda Hodkiewicz [c]

[a] *Department of Computer Science and Software Engineering, University of Western Australia, Crawley, 6009, Australia*
*E-mails: caitlin.woods@research.uwa.edu.au, tyler.bikaun@research.uwa.edu.au*
[b] *Industrial AI Research Centre, UniSA STEM, University of South Australia, Adelaide, 5000, SA, Australia*
*E-mails: matt.selway@unisa.edu.au, mst@cs.unisa.edu.au*
[c] *School of Engineering University of Western Australia, Crawley, 6009, Australia*
*E-mail: melinda.hodkiewicz@uwa.edu.au*

**Abstract.** Maintenance of assets is a multi-million dollar cost each year for asset intensive organisations in the defence, manufacturing, resource and infrastructure sectors. These costs are tracked though maintenance work order (MWO) records. MWO records contain structured data for dates, costs, and asset identification and unstructured text describing the work required, for example 'replace leaking pump'. Our focus in this paper is on data quality for maintenance activity terms in MWO records (e.g. replace, repair, adjust and inspect).

We present two contributions in this paper. First, we propose a reference ontology for maintenance activity terms. We use natural language processing to identify seven core maintenance activity terms and their synonyms from 800,000 MWOs. We provide elucidations for these seven terms. Second, we demonstrate use of the reference ontology in an application-level ontology using an industrial use case. The end-to-end NLP-ontology pipeline identifies data quality issues with 55% of the MWO records for a centrifugal pump over 8 years. For the 33% of records where a verb was not provided in the unstructured text, the ontology can infer a relevant activity class.

The selection of the maintenance activity terms is informed by the ISO 14224 and ISO 15926-4 standards and conforms to ISO/IEC 21838-2 Basic Formal Ontology (BFO). The reference and application ontologies presented here provide an example for how industrial organisations can augment their maintenance work management processes with ontological workflows to improve data quality.

Keywords: Maintenance work order, ontology, natural language processing, centrifugal pump

## 1. Introduction

Maintenance of assets is a significant cost input for the manufacturing, resources, defence and infrastructure sectors. Maintenance costs typically range between 20–60% of operational expenditure depending on industry and asset

*Corresponding author. E-mail: caitlin.woods@research.uwa.edu.au.

type [60]. There has been significant effort in the last decade to move from reactive to preventative and predictive maintenance strategies propelled by developments in sensing, WiFi, cloud computing, and data analytics. However, generating value from analytics using these platforms has often proved challenging [1,6,67]. In part, this is due to the way in which data describing what maintenance work was actually done, to what item, when it was done and what it cost, is captured and stored [9,56] in maintenance work order (MWO) records; that is, as unstructured free text.

MWOs are the equipment equivalent of health records for humans. Tens to hundreds of thousands of MWOs are generated a year on a moderately complex site or asset system. A MWO is an information artifact that is generated in industrial organisations (that are mature enough to maintain a maintenance strategy) to inform technicians that work needs to be done. MWOs can be generated automatically (by a Computerised Maintenance Management System) in the case of preventative maintenance, or manually by a technician in response to an observation or event. These MWO records contain an unstructured text field with a short description of the job (e.g., "replace all damaged idMobile lers" or "inspect / refurb scrapers"). Each record also has structured data fields for a unique work order number, asset functional location, dates for the proposed and actual start and end date of the work, budget and actual costs, as well as fields for metadata about the record such as when it was generated and closed. Computer readability of the unstructured text in MWOs using natural language processing (NLP) is currently an active area of research and development [7,12].

Our focus in this paper is on data quality of the maintenance activity term in the MWO unstructured text field. This activity is described by verbs such as replace, repair, adjust and inspect. This verb is written manually into the unstructured short text of the record by an operator or maintainer. If this information is unclear or unreliable it adds uncertainty to the identification of life cycle events such as end-of-life and when preventative actions were done (or not). While it is desirable to predict a failure event from condition monitoring data, the engineer also needs to know if action was taken based on the prediction, what the maintenance activity was, and if the action was taken before the failure manifest. The raw data from which this action can be inferred is stored as structured and unstructured text in maintenance work order records (MWOs). Currently this inference is done manually by engineers and planners.

Data quality of MWOs has been a topic of organisational behaviour research for many years [37,40,64]. However the problem is unresolved and data quality is cited as an issue affecting maintenance and warranty management improvement programs [18,33,47]. Labour productivity in maintenance is now a focus of senior management attention [60]. The time taken by experienced maintenance planners and reliability engineers to manually read and process MWOs to get information for their analysis is a significant impediment to improving their productivity [15]. Machine augmentation of this basic task is required.

Our interest is improving data quality using ontology-based reasoning on MWO unstructured texts pre-processed with natural language processing techniques. Currently questions such as 'was pump 101 actually replaced on 12/1/20?' are resolved manually by a reliability engineer [10,54]. Such analysis requires the reliability engineer to examine MWO records one at a time as described in [3]. However, the reasoning used to answer a number of these questions can be formalised into a set of rules [3]. This opens the door to the use of ontologies as part of the MWO data quality improvement process. The specific focus of this paper is an ontology for maintenance activity terms.

We propose a solution to improve MWO data quality using a reference ontology and an application-level ontology. The reference ontology contains a holistic view of the types of activities conducted by maintenance personnel. Elucidations based on ontological analysis are provided for seven core maintenance activity terms. The application-level ontology uses the reference ontology and real-world MWO data to perform the necessary reasoning tasks for our use case. This ontology examines each MWO and determines if the maintenance activity word used in the MWO is consistent with the other information in the maintenance record.

If the activity described in a MWO's unstructured text matches the record's structured data, then we can be more certain that the activity as described actually occurred in practice. This information can then be used, with confidence, for a number of important downstream analytics tasks such as in calculating reliability metrics such as mean-time-to-failure, in root cause analysis following major undesirable events, and for assessing the effectiveness of maintenance strategy [33,42,43,46].

The paper is organised as follows. Section 2 describes previous ontology development and industry reference data models in the maintenance area and provides details on MWOs and their contents. The use case, including instance data, is presented in Section 3. Section 4 describes steps in the process of identifying seven maintenance

activity classes and their elucidations. Ontology design and competency questions are presented in Section 5 and the performance of the ontology is assessed in the evaluation in Section 6. Finally, Section 7 critically examines the performance of the ontology and identifies opportunities for further work.

## 2. Background

### 2.1. Maintenance

Maintenance is defined by an International Standard as "the actions intended to retain an item in, or restore it to, a state in which it can perform a required function" [20]. We note from this that the notion of an action or activity is central to the concept of maintenance.

In mature maintenance organisations, each maintainable item has an associated maintenance strategy. The maintenance strategy determines what work should be done, when to do it and at what level in the asset hierarchy it should be performed. Deciding on maintenance strategy is a well established process based on failure modes and effects analysis and reliability centred maintenance (RCM). These are described in international standards such as [21,52]. The 'what strategy?' decision is informed by factors such as the safety, environmental, production or cost consequences of an item's failure, whether loss of function is technically and cost effective to observe, and how long it takes from observation of deterioration to failure event. RCM decision-logic guides the user towards one of the following strategies: a) use based, also known as fixed interval restoration/repair/inspect strategy, b) strategy based on condition monitoring and inspections, c) failure-finding and d) run-to-failure [17]. Colloquially, these are classified into preventative and corrective maintenance strategies, with *preventative* including fixed interval work and periodic inspections such as condition monitoring and *corrective* covering failure finding, run-to-failure, and work arising from preventative tasks. The notions of preventative and corrective strategies are relevant to the reasoning used in this paper.

### 2.2. Review of previous ontology work on maintenance activity terms

There is increasing interest in automating tedious data processing tasks that require the input of experienced people and in being able to automatically process equipment maintenance and failure data using ontologies for decision support to standardise the decisions made when analysing the data [15,33,42,46]. Interest is high in industries such as defence, aerospace, oil and gas, and manufacturing and this is compelling groups to come together to consider standardisation as they recognise the scale of the task and the challenge of going it alone. In 2021 the ISO/IEC 21838-1 for Top Level Ontologies [26] was issued alongside ISO/IEC 21838-2 Standard [27] describing Basic Formal Ontology (BFO). Standardisation of a further two top level ontologies, Descriptive ontology for linguistic and cognitive engineering (DOLCE) and TUpper is in progress. This standardisation process is a key plank in industry acceptance of ontologies. There is a surge of activity to develop industry-relevant, domain specific ontologies, that are aligned to one of these Top Level Ontologies. Of relevance to this paper is the work of the Industrial Ontology Foundry (IOF) and their work to publish a domain ontology for manufacturing aligned to BFO. The IOF promotes a principles-based approach to the design of ontologies for design, maintenance, supply chain, production, and lifecycle management of equipment. A number of the terms and relations necessary to support description of maintenance activities are already formalised in the IOF [59].

While there have been a number of papers published on maintenance ontologies these have, in the main, been developed using a top down approach. Some have tried to capture maintenance in general [30,35] while others have focused more specifically on various processes in maintenance such as maintenance work management [19], failure modes and effects analysis [16], fault classification from warranty records [48], modelling maintenance strategies [38], and in criticality assessment [44]. Dealing with real-life maintenance data has not been the primary motivation for ontology development. Instead, conceptual models have dominated with limited use cases selected to support and illustrate specific reasoning and conceptualisation challenges. There have been some exceptions. For example, in the automotive industry, Rajpathak and General Motors have used natural language processing and ontologies to extract data from hundred of thousands of unstructured warranty records [46,48,49] and there is similar interest in

Table 1

Examples of other fields of interest in MWO records

| Field(s) | Description | Field type |
|---|---|---|
| Functional location | Asset location where maintenance is performed | Code |
| Work description | Description of the problem or work to be done | Unstructured text |
| Work order number | Unique identifier for the record | Numeric |
| Work order type | Code to describe if the work is corrective or preventative | Code |
| Start and end dates (MWO) | Metadata for the MWO record | Date |
| Scheduled start and end dates | Scheduled start and end date for the proposed maintenance activity | Date |
| Actual start and end dates | Actual start and end date for the proposed maintenance activity | Date |
| Estimated and actual labour hours | Budget and actual manpower hours for the maintenance activity | Numeric |
| Estimated and actual labour costs | Budget and actual manpower cost for the maintenance activity | Numeric |
| Actual parts costs | Cost of parts used in the maintenance activity | Numeric |

the analysis of safety data from accident databases [57]. In the oil and gas industry, there is emerging work on using ontology patterns to ingest instance data from MWOs and failure modes and effects analysis records into ontologies for use in reasoning about the effectiveness of maintenance strategy for a complex operating asset [32].

When looking for previous ontology work specifically on maintenance activity classes we located a list of 16 drawn from a data set of 654 activities/tasks for a wire harness assembly case study [41]. However, these activities are defined using natural language phrases such as 'laying cable flat' and 'inserting into the tube or sleeve' and can be used only in their specific case study. The activity classes identified are not generic to maintenance in general.

## 2.3. Maintenance work order records

A MWO record is generated every time work is done, or needs to be done, on an asset. MWO records are the only digital record of maintenance events for the life history of an asset. Hundreds of thousands of these MWOs are generated every year in asset intensive organisations. In a MWO record the primary interest to the reliability engineers are the fields described in Table 1. The functional location is the asset identifier, the work order description contains the unstructured text of interest in our work, the work order type identifies if the work is preventative or corrective, and there are fields for the start and end dates and various budget and actual costs. Other fields, not shown, include metadata for the record. The information contained in MWOs is the only record of what maintenance work is done when, to what item, and how much it cost.

## 2.4. Natural language processing of maintenance work orders

The reader will note from Table 1 that there is no dedicated fields for key information such as "what specific item needs work?", "what work needs to be done?", "why does it need to be done?". The answers to these questions have to be inferred from a 4–8 word sentence captured in the 'Work description' field. A typical example, drawn from [13] is 'change out leaking engine'. This phrase is illustrative of MWOs in general in that it contains an activity word 'change out', a state or problem word 'leaking' and an item identifier 'engine'. Work to develop pipelines to separate the words in these MWOs into classes using natural language processing is very active [7,13,42,46,55,62]. Once extracted, the information in these activity, item, and state fields is being used for visualisation and analysis to support a number of maintenance and reliability decisions [3,48,53,54]. However, one challenge for all users is the large number of terms in all these classes resulting from synonyms, misspellings, different tenses, abbreviations, and jargon.

Work on defining classes for types of maintenance *activity* in ontology is limited. There is a list of 10 maintenance action codes (inspection, servicing, adjustment, alignment, resetting, replenishing, coating, repair, replace, overhaul) produced by [28] but the selections on the list are not justified nor the terms fully described. A Bag-Of-Words approach on a corpus of 690,000 MWOs identified 103 maintenance terms but acknowledged that many terms such as service that can be a noun or a verb were missed [13]. There was no attempt to group these terms into manageable groups. The raw data arising from these works are not freely available and the resulting lists of classes are not clearly defined. So instead we turn to lists available in the engineering standards literature.

Table 2

Maintenance activity terms and associated descriptions from ISO14224 [23] and ISO 15926-4 [24]

| Maintenance activity term | ISO 14224 descriptions. The (C, P) indicate activities performed under corrective and/or preventative maintenance strategies. | ISO 15926-4 descriptions |
|---|---|---|
| Replace | Replacement of an item by a new or refurbished item of the same type and make (C, P). | Acting by putting something new in the place of. |
| Repair | Manual maintenance action performed to restore an item to its original appearance or state (C). | Restoring by bringing an item back in its original state (shape and properties) or in a revised design state. |
| Modify | Replace, renew or change the item, or a part of it, with an item/part of a different type, make, material or design (C, P). | Acting by replacing, renewing, or changing items, or parts with another item or part of different type, make, material or design. |
| Adjust | Bringing any out of tolerance condition into tolerance (C, P). | Changing by bringing in a more satisfactory state. |
| Refit | Minor repair/servicing activity to bring back an item to an acceptable appearance, internal or external (C, P). | |
| Check | The cause of failure is investigated but no maintenance action is performed, or action is deferred. Able to restart by simple actions e.g.restart or resetting (C). | Verifying by investigating and obtaining confirmation or substantiation of accuracy, fitness, or due performance |
| Service | Periodic service tasks: normally not dismantling of the item (e.g. cleaning, oil top up etc) (P). | (Definition not relevant to maintenance.) |
| Test | Periodic test of function or performance (P). | Validating that a physical object satisfies specified criteria; typically by verification of a property value of the physical object. |
| Inspect | Periodic inspection/check: a careful scrutiny of an item carried out with or without dismantling, normally by the use of senses. (P). | Acting by viewing closely and critically in order to ascertain quality or state, detect errors, or otherwise appraise |
| Overhaul | Major overhaul (C, P) such as a comprehensive inspection/overhaul with extensive disassembly and replacement of items as specified or required. | Repairing as to restore to satisfactory working order. |
| Calibrate | | Adjusting precisely for a particular function, or standardising by determining the deviation from standard and adjusting if appropriate. |
| Maintain | | Keeping in a state of repair, efficiency, or validity, preserving from failure. |
| Troubleshoot | | Maintaining by locating trouble and making repairs in machinery and technical equipment. |
| Renew | | Revising by replacing old parts with new ones. |

### 2.5. Lists of maintenance activity terms from international standards

In this work we draw on two international standards, ISO 14224 (2016) and ISO 15926-4, for their maintenance activity terms and definitions. The terms, definitions, and their alignment between the two standards is shown in Table 2. The following sections briefly introduce the two standards.

### 2.5.1. ISO 14224

The ISO 14224 standard (Petroleum, petrochemical and natural gas industries – Collection and exchange of reliability and maintenance data for equipment) [23] is an influential international standard originally developed for the oil and gas industry but now widely used in other process and heavy industry sectors. This standard was originally developed in the early 1980s due to work by the oil and gas sector seeking to standardise maintenance and failure data collection for an Offshore Reliability Data handbook called OREDA [58]. Data quality is a key consideration in OREDA because the data in this handbook is so widely used by industry for safety-critical design decisions [50]. ISO 14224 has been through 5 editions and contains widely-used lists for equipment hierarchies, failure modes, mechanisms and causes, and a set of maintenance activity terms. We list the activity terms and their definitions from ISO 14224 in Table 2.

Table 3

Mapping of activity super-class terms from ISO/TS 15926-4 (industrial automation systems and integration – integration of life-cycle data for process plants including oil and gas production facilities – part 4: initial reference data) showing their maintenance-related subclass terms

| Activity super-class | Text description (ISO 15926-4) | Maintenance-related subclasses from ISO 15926-4 |
| --- | --- | --- |
| acting | Activity of carrying out a process of change or alteration | replacing → acting<br>overhauling → repairing → restoring → putting → acting<br>renewing → revising → improving → enhancing →<br>increasing → making → acting<br>testing → validating → providing → supplying → acting<br>service → service-acting → acting<br>inspecting → acting<br>cleaning → removing → acting<br>calibrating → adjusting → changing → making → acting<br>modifying → acting<br>analyzing → studying → acting<br>greasing → lubricating → applying → using → acting |
| being | Activity that entails the having of an objective existence | troubleshooting → maintaining → keeping →<br>holding-having → having → being |

### 2.5.2. ISO 15926-4 activity class hierarchy

The ISO/TS 15926-4 Standard (Industrial automation systems and integration – Integration of life-cycle data for process plants including oil and gas production facilities – Part 4: Initial reference data) is a data model developed for the engineering design community particularly in the process sector. It includes a number of reference data models for activities, rotating equipment, static equipment, instrumentation, units of measure, valves etc. In the activity data model there are 1795 activity terms organised into a deep taxonomy (7 levels) from a superclass set of 22 terms. These terms are: acting, absorbing, affirming, behaving, being, choosing, competing, consigning, directing, disjoining, drawing, happening, occurring, passing-letting, process, pushing, reacting, reducing, refusing, relieving, squeezing, watching. Maintenance activity terms relating specifically to interactions with equipment such as replacing, repairing, inspecting sit under the superclass *acting* with the exception of troubleshooting and maintaining which are classified under *being*. A summary of the maintenance-related terms and their subclass structure is shown in Table 3.

### 2.6. Distinguishing between maintenance activity terms

The descriptions for maintenance activities drawn from ISO 14224 and ISO 15926-4 are outlined in Table 2. In these descriptions we can see there are two motivations for a maintenance activity. The first is to 'change (or restore) the state' of the item. The second is a 'discovery' process to determine if work is required. Under the 'change of state' notion, either: a) the state of the item is known to be deteriorated and needs restoration, e.g., for repair, adjust, and corrective replacement; or b) work is required as a result of pre-determined strategy at a predetermined interval regardless of the state of the item, e.g., preventative replacement, fixed interval calibration, test and service. Under the 'discovery' notion, the state of the item is unknown, but a problem is suspected. The state of the item needs to be determined, e.g., by a diagnostic activity such as a test or inspection.

However, we note that these two standards do not agree on several of the term definitions or the same set of terms. For example, the terms calibrate, maintain, troubleshoot and renew appear in ISO 15926-4 but not in ISO 14224. Conversely, the terms refit, service appear in ISO 14224 but not ISO 15926-4. The term overhaul in ISO 14224 is superficially defined as an 'overhaul' and in ISO 15926-4 as 'repairing or restore to satisfactory working order' which is only distinguished from repair – defined in ISO 15926-4 as 'bringing an item back in its original state or in a revised design state' – by someone being able to separate between an item in 'satisfactory working order' or 'original or revised design state'. What is meant by a 'revised design state' is not clear. The definition for refit is only in ISO 14224, and described as 'bringing back acceptable appearance' rather than a function, which adds another level of ambiguity by the introduction of the word 'appearance'.

To move forward in the presence of this partial agreement on terms and their definitions in the ISO 14224 and 15926-4 Standards, we have captured general concepts in Table 2 in a the classification workflow shown in Fig. 1. In
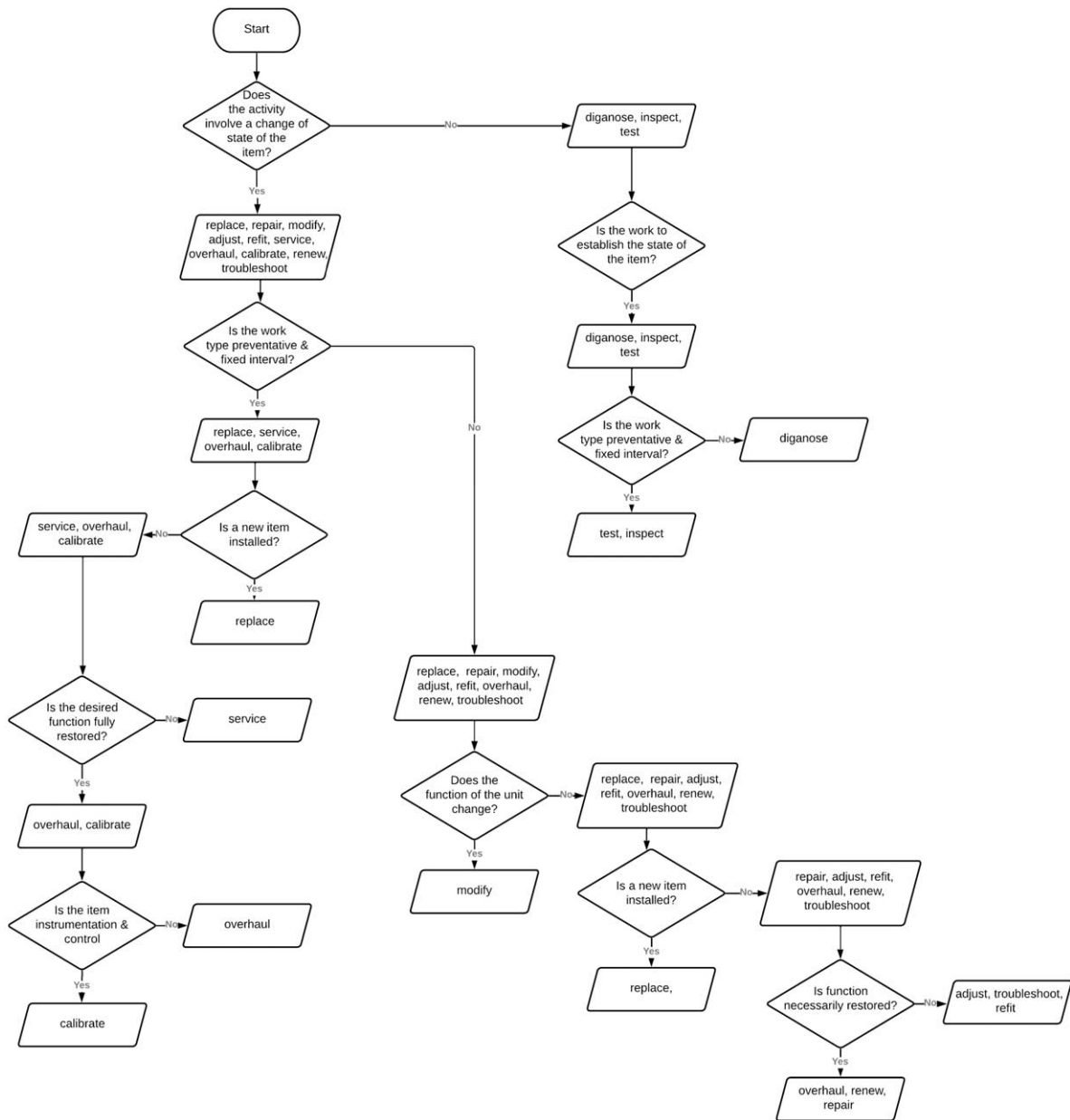
Fig. 1. Classification workflow capturing logic to distinguish between maintenance activity terms.

this workflow, diamonds contain questions for the users to answer and depending on the yes/no answer, an inference about the nature of the activity is shown in a parallelogram. In developing this we focus on information of potential value to reasoning as follows:

– Is the activity initiated by a preventative maintenance strategy or is it corrective maintenance?
– Is the item performing its desired function?
– Does the activity involve an action that restores function?
– Does the activity change the function and/or capability of the item?
– Is the state of the item known or unknown?

Table 4

Equipment subdivision for a centrifugal pump-motor system into subunits and maintainable items

| Subunit | Driver and electrical | Power transmission | Pump unit | Control and monitoring | Lubrication system | Piping and valves |
|---|---|---|---|---|---|---|
| Maintainable items | Motor<br>Variable drive<br>Power supply | Gearbox<br>Coupling | Pump<br>Casing<br>Impeller<br>Shaft<br>Bearing<br>Mechanical<br>Seal<br>Baseplate<br>Packing | Actuator<br>Controller<br>Sensor<br>Pressure switch<br>Flow switch<br>Flowmeter | Grease<br>Oil<br>Breather<br>Filter | Pipe/Piping<br>Valve<br>Check valve |

## 3. Use case

A corpus of +800,000 unstructured MWO descriptions from rail, process and mining sectors was processed using NLP to identify activity words used by maintenance and operations personnel (this is described in Section 4.1.1). The use case for demonstrating reasoning using this ontology is real-world industrial data for centrifugal pump system in a process plant. The general subunit breakdown of a centrifugal pump system is given in Table 4 while Table 5 contains a set of MWOs for a single functional location for a single pump in a process plant from 2012–2020. There are hundreds of other similar pumps in this facility. Columns 2–3 and 6–8 in Table 5 are representative of a data set a reliability engineer would normally work with. These fields are usually of most interest to the reliability engineer and are often the most readily available as reports downloadable as CSV files from the computerised maintenance management system (CMMS). The first column contains an ID to assist us with discussions of the data set and the analysis thereof. The second column contains a date, in this case we have selected the date on which the work was started. The third column contains the unstructured text describing the work and is subject to natural language processing. The sixth column describes the work order type as either corrective or preventative and the final two columns contain the labour and material costs; referring to Table 1, only one of labour hours or labour costs is required, in this case the records include cost, while material cost is equivalent to the actual parts cost.

Table 5 contains two columns not found in the CMMS. The fourth and fifth are calculated columns that capture the outputs of processing the unstructured text with an NLP pipeline. The ability to do this depends on lexical normalisation [63] and annotation of a large corpus of MWOs [62] from fixed and mobile plant assets which has been used to train a deep learning model. The outputs from this processing that are of interest to the work in this paper are the extraction of the *item* and *activity* in the unstructured text. For each NLP-identified *item*, we have also provided the item's subunit. This subunit has been extracted from Table 4. The NLP pipeline that enables the work presented in this paper is discussed in further detail in Section 4.1.1.

Over the 7.5 year period there have been 36 MWOs raised against the pumping system in this functional location. The majority (16 out of 36) are related to the pressure switch. The pressure switch is part of the Control and Monitoring subunit. The pump has been replaced once, in 2017, at a material cost of > \$15,000. A costly activity was performed on the pump housing due to an oil leak in 2018 for ≈ \$8,000.

The MWOs include both preventative and corrective work, with preventative work being associated with a semi-structured form in the unstructured text field usually starting with some periodic interval such as '78W' for 78 weeks. Preventative work passes through the planning and scheduling process in the maintenance work management system [31]. Corrective work orders are urgent and need to be actioned immediately by shift maintainers, whereas preventative works go into the maintenance management process to be dealt with by maintenance planners and then scheduled into a weekly maintenance plan. In general, there is a desire to move towards maintainers spending their time on preventative work with correctives generated from preventative inspections. Correctives are costly and disruptive and typically represent a failure of strategy, except where they associate with a deliberate run-to-failure strategy.

Finally, of note is that some MWOs (e.g., 6, 12, and 19 in Table 5) have \$0 labour cost, which is interpreted by reliability engineers as unexecuted work. This frequently occurs as work can be scheduled but not actioned due to

Table 5

Example of a data set of maintenance work orders for a single pump over the period of 2012–2020. Abbreviations used: D& E – driver and electrical, PT – power transmission, PU – pump unit, C&M – control and monitoring, LS – lubrication system, P&V – piping and valves

| ID | Date | Unstructured (raw) text | NLP-identified item (subunit) | NLP-identified activity | Work order type | Labour cost ($) | Material cost ($) |
|---|---|---|---|---|---|---|---|
| 1 | 2012-10-18 | pressure switch undersize | pressure switch (C&M) | none | corrective | 560 | 821 |
| 2 | 2013-02-14 | remove pressure switch | pressure switch (C&M) | remove | corrective | 1292 | 711 |
| 3 | 2013-03-19 | calibrate pressure switch | pressure switch (C&M) | calibrate | corrective | 413 | 0 |
| 4 | 2013-04-29 | pump not pumping well | pump (PU) | none | corrective | 516 | 0 |
| 5 | 2013-05-21 | pressure switch leaking | pressure switch (C&M) | none | corrective | 560 | 700 |
| 6 | 2013-06-03 | calibrate pressure switch | pressure switch (C&M) | calibrate | corrective | 0 | 0 |
| 7 | 2013-06-18 | 18M electrical service motor | motor (D&E) | service | preventative | 258 | 0 |
| 8 | 2013-06-28 | calibrate pressure switch | pressure switch (C&M) | calibrate | corrective | 322 | 0 |
| 9 | 2013-08-07 | calibrate pressure switch | pressure switch (C&M) | calibrate | corrective | 400 | 0 |
| 10 | 2013-10-06 | calibrate pressure switch | pressure switch (C&M) | calibrate | corrective | 413 | 1455 |
| 11 | 2013-10-06 | valve needs replaced | valve (P&V) | replace | corrective | 650 | 344 |
| 12 | 2013-12-23 | investigate oil leak | oil (LS) | investigate | corrective | 0 | 0 |
| 13 | 2014-02-10 | seal leaking | mechanical seal (PU) | none | corrective | 3195 | 5062 |
| 14 | 2014-03-20 | pressure switch unserviceable | pressure switch (C&M) | none | corrective | 680 | 1133 |
| 15 | 2014-05-31 | check pressure switch | pressure switch (C&M) | check | corrective | 145 | 0 |
| 16 | 2014-07-01 | pump not pumping | pump (PU) | none | corrective | 64 | 0 |
| 17 | 2014-11-04 | install new pressure switch | pressure switch (C&M) | install | corrective | 400 | 0 |
| 18 | 2014-12-16 | 78W electrical service motor | motor (D&E) | service | preventative | 131 | 0 |
| 19 | 2014-12-31 | 26W mech service pump | pump (PU) | service | preventative | 0 | 0 |
| 20 | 2015-03-25 | pressure switch failure | pressure switch (C&M) | none | corrective | 82 | 0 |
| 21 | 2015-04-20 | replace pressure switch | pressure switch (C&M) | replace | corrective | 348 | 1200 |
| 22 | 2015-05-16 | motor tripping on high amps | motor (D&E) | none | corrective | 270 | 0 |
| 23 | 2015-05-19 | repair pressure switch | pressure switch (C&M) | repair | corrective | 541 | 0 |
| 24 | 2015-06-05 | 26W mech service pump | pump (PU) | service | preventative | 322 | 110 |
| 25 | 2015-11-24 | 26W mech service pump | pump (PU) | service | preventative | 478 | 110 |
| 26 | 2016-06-17 | 78W electrical service motor | motor (D&E) | service | preventative | 91 | 0 |
| 27 | 2017-06-04 | replace pump | pump (PU) | replace | corrective | 5168 | 15370 |
| 28 | 2017-06-12 | 78W electrical service motor | motor (D&E) | service | preventative | 868 | 0 |
| 29 | 2018-06-15 | change oil | oil (LU) | change | corrective | 601 | 0 |
| 30 | 2018-07-03 | oil leak from housing seal | mechanical seal (PU) | none | corrective | 8632 | 8369 |
| 31 | 2019-01-06 | pump is tripping | pump (PU) | none | corrective | 282 | 0 |
| 32 | 2019-06-06 | 78W electrical service motor | motor (D&E) | service | preventative | 726 | 0 |
| 33 | 2019-11-26 | faulty flowmeter | flowmeter (C&M) | none | corrective | 269 | 0 |
| 34 | 2020-02-19 | repair pump | pump (PI) | repair | corrective | 3207 | 566 |
| 35 | 2020-04-14 | pressure switch faulty | pressure switch (C&M) | none | corrective | 99 | 0 |
| 36 | 2020-04-22 | replace pressure switch | pressure switch (C&M) | replace | corrective | 198 | 2021 |

production, resourcing, or other issues. Consequently, it is important to be track work that is not executed and we use this information in our reasoning.

There are many data models to describe how asset systems can be decomposed into functional or physical breakdown. A detailed review of these is beyond the scope of this paper. Our taxonomy is informed by the hierarchy in ISO 14224 [23] of maintainable item, subunit, and equipment unit. The maintainable items have been separated into a number of subunits as shown in Table 4. The proposed structure mirrors one commonly found in asset registers and computerised maintenance management systems in our industry partners. We use the information on where maintainable items sit in the hierarchy to support reasoning in the activity ontology.

## 4. Reference ontology for maintenance activities

In this section, we present a reference ontology containing the types of activities that are typically performed in maintenance. This reference ontology contains elucidations for seven core maintenance activities. Given the inconsistencies in the existing standards presented in Section 2.6, we decided to extract reference-level activity terms from NLP analysis of +800,000 real-world MWO records. We then use the existing standards to distinguish between these terms and inform their elucidations. The methods that we use to extract the activity terms and group them according to seven core activities is described in Section 4.1. Elucidations for these activity terms, and the ontological choices underpinning them, are provided in Section 4.2. The purpose of the reference ontology is to provide a set of terms and loose definitions to be used in future ontological analysis by the wider research community. We demonstrate the use of this reference ontology as an input to our application-level model described in Section 5.

### 4.1. Developing notions for maintenance activity terms

#### 4.1.1. Step 1: Activity term identification

To generate a holistic view of the types of activities performed by maintenance technicians, we used a corpus of +800,000 maintenance work orders from heavy mobile equipment, fixed plant, and rail assets. The work order data is similar to the sample data shown in Table 5. Initially we explored using off-the-shelf part-of-speech (POS) taggers (NLTK and spaCy) to extract verbs automatically. However, these approaches could not extract verbs in most cases due, in part, to the informal grammar inherent in these short maintenance work descriptions [7,12]. There are currently no off-the-shelf pre-trained/configured POS algorithms for maintenance work orders. To overcome this limitation, we applied named entity recognition to our corpus using a pre-trained neural model fine-tuned on annotated maintenance work descriptions extracted from these asset types. This pre-trained entity recognition model created by the UWA NLP-TLP group [66] is the product of 3 years research effort involving schema development [13], lexical normalization [2], and collaborative annotation [62] of MWOs. A description of the end-to-end pipeline is available [61]. The entity recognition model is fine-tuned on a set 6,000 high-quality, gold-standard, annotated MWOs and extracts notions such as *items*, *activities* and *states* from the text.

The processing steps are described below.

- Apply the pre-trained NER model to our corpus of +800,000 unstructured maintenance work order, filtering extracted classes to tokens tagged as 'Activity' (i.e. activity terms).
- Produce a list of more that 470,000 activity terms including duplicates, synonyms and misspellings, of this 10,860 were unique terms.
- Perform frequency analysis over the extracted activity terms to identify high-frequency occurrences.
- Filter the activity terms to the top 99% by a frequency that constituted our primary collection of terms. This resulted in a list of 109 terms that represent 99% of the usage due to the very long tail of low frequency words. In the final list of 109 terms the top two terms by frequency of occurrence (replace and repair) represent 41% (470,000) of the activity terms extracted.
- Examine the resulting list of 108 terms, manually identify misspelling and run a script to 1) normalise words such as replaced, replacing, repl, relace, replacement to replace, 2) change past tense to present, and 3) remove the 're' in reweld, re-seal, resecure and so on to the weld, seal and secure, for example. The resulting 108 terms are discussed below.

#### 4.1.2. Step 2: Synonym grouping

A number of high frequency use activity terms are immediately apparent on examination of the NLP pipeline results. Replace occurs more than 120,067 times in the corpus of +800,000 activity terms. Likewise the terms repair (73,340), inspect (21,478), service (22,245), check (18,796) and adjust (5,303). These six activity terms map well onto the maintenance activity terms identified by both ISO 14224 and ISO 15926-4 in Table 2.

The team used subject matter experts to group the remaining words to maintenance activity classes described in the paragraph above which align with terms proposed by ISO 14224 and ISO 15926-4 shown in Table 2. In performing this synonym grouping consideration was given to a) the semantics of the activity word, b) mention

Table 6

Maintenance activity terms and semantic variants identified using NLP from the top 1% of words by frequency of occurrence (shown in []) in the +800,000 corpus

| Maintenance activity term | Synonyms from the NLP analysis. [] shows frequency of occurrence in corpus | Corrective (C) or Preventative (P) |
|---|---|---|
| Replace | replace [120,067], change-out [30,721], fit [17,694], remove [7253], install [4505], connect [414] | (C, P) |
| Repair | repair [73,340], seal [7,706], weld [2,478], overhaul [2,335], rectify [1,184], rebuild [1,044], mount [745], torque [702], refurbish [673], fix [629], build up [568], rewheel [351], rewire [350], renew [294], heat [267], rerail [240], regas [240], grind [218] | (C) |
| Inspect | inspect [21,478], monitor [1,176], ndt [1,093], measure [416], crack test [206], | (P) |
| Adjust | adjust [5,303], tighten [1,135], position [573], tune [539], straighten [428], turn [380], set [340], shim [340], top up [275], tension [247], tilt [246], regulate [167] | (C) |
| Service | service [22,245], clean [3,412], fill [1,242], sample [1,198], charge [660], rotate [459], drain [421], wash [407], lubricate [385], grease [245] | (P) |
| Diagnose | check [18,796], investigate [3,882], test [2,875], diagnose [1,959], check out [573], analyse [192] | (C) |
| Calibrate | calibrate [251] | (P) |
| Supporting activities | report [3,359], modify [2,095], fabricate [2,066], dispatch [864], start [843], start [843], jump start [840], audit [797], make up [787], return [774], assist [704], purchase [692], upgrade [680], implement [602], make [602], secure [557], perform [516], quote [487], set up [484], contract [472], access [450], move [440], find [411], complete [370], add [340], prepare [308], manufacture [286], support [283], pick up [276], commission [257], training [247], reroute [242], load [242], action [231], review [227], isolate [223]. disconnect [213], shutdown [209], unload [208], cost [199], rekit [197], transfer [192], release [180] | Not relevant |

of the class as part of maintenance strategy development as described by reliability-centred maintenance (RCM) philosophy and standards [39,52], and c) associating the verb with either preventative or corrective work. This work produced eight synonym groups, seven maintenance-related activity terms and a group associated with supporting activities. The resulting list of maintenance activity terms and their semantic variants is presented in Table 6.

We propose calibrate as a top-level concept. Calibration is central to maintenance and involves the organised testing and adjustment of instrumentation as part of preventative maintenance strategy. Calibration is done by control and instrumentation technicians and is also defined subclass in an asset hierarchy [23] and thus can be used in the reasoning. Sensors and automation are a key part of the future of managing assets and hence calibration is likely to be much larger part of maintenance work and work orders than it is in the historical set we have drawn on for our NLP analysis.

We propose diagnose as the label for the grouping of terms containing check, investigate, test, diagnose, check out and analyse. Diagnose conveys a broader set of activities than the notion of check which is more primitive step in an investigative process. The need for maintainers to perform diagnostic work is also increasing with the increased use of condition based and predictive maintenance strategies [8].

We consider overhaul a synonym to repair. As can be seen in the definitions of overhaul in Table 2 there is no clear distinction between an overhaul and a repair from the activity alone. Instead, the distinction comes from what is being worked on (i.e. work at the equipment or at the part level). This distinction belongs outside of an ontology designed for describing activities as it depends on several factors that are specific to individual organisations. For example, remote operations may not have the people or equipment to overhaul on site, warranty arrangements might require that overhaul activities be done by the original equipment manufacturer and not by the organisation's maintenance team, and so on. In practice the choice of the term overhaul or repair is dependent on the site and industry cultural norms.

We have removed refit and have grouped this word under repair as there are no established rules to distinguish between them other than refit is a small repair.

We have moved modify to the *supporting activity* terms. In Moubray's classic text on reliability-centred maintenance [39], modify is considered a synonym of redesign. Redesign is "any change to the specification of a component". Moubray deals with directly with the question 'is redesign part of design or maintenance?' He argues that redesign is part of design as modifications can take from 6 months to 3 years from conception to commissioning whereas "the maintenance person has to maintain the equipment as it is today, not what should be there or what might be sometime in the future" [39]. In the international standard for RCM, modification is not a maintenance task coded as part of maintenance strategy [51]. Similarly, ISO 14224 indicates that modification is not strictly a maintenance activity but may be performed by people with maintenance capabilities [23].

The activities in supporting activities are not the primary focus on this work but we have made an initial set of activities into which the verbs in Table 6 have been grouped. These are admin, assemble, isolate, modify, move, operate, perform, teamwork.

Combining the grouping of NLP extracted terms with the standards literature identified a set of core maintenance activity types for this ontology. In addition, we identified distinctions between these activity types that, in practice, have not been captured in the definitions in the standards. However, having identified these seven core concepts, the label used to describe the concept can be adjusted to suit specific organisations through edits to the annotations.

### 4.1.3. Step 3: Descriptions of the maintenance activity terms

Table 7 provides a subject matter expert natural language description, a semi-formal description and an elucidation each of the seven maintenance activity terms. The elucidations refer to concepts defined by BFO, IOF Core, and maintenance state ontology [68] using *italics*. We describe the ontological choices underpinning these elucidations in Section 4.2. The elucidations only state necessary, but not sufficient, conditions for the concepts. Terms are defined in the absence of information that would ideally be modelled ontologically but generally is not captured in practice. Hence, as will be discussed later, the need for rules based on expert knowledge to perform the classification of activities.

### 4.2. Discussion on elucidations for the core activity terms

In this section, we discuss the elucidations given in Table 7 and the ontological choices underpinning these elucidations. These elucidations draw on our previous work on Maintenance State to help distinguish between activity types [68]. This work on maintenance state describes terms that are used in these elucidations such as operating state.

First, we describe the elucidation for the replace activity. This elucidation describes a situation where there are two material artifacts (**a** and **b**) that have functions of the same type, **F**. When the replace activity occurs, **a** is a continuant part of a system before the replacement, and **b** is a continuant part of the system after the replacement. This is true whether a whole piece of equipment is replaced (i.e. a pump) or only part of a system is replaced (i.e. a pressure switch). A critical part of this description is that both the item to be replaced (**a**) and the item to be installed (**b**) both fulfil the same *function* when installed as part of a system. In other emerging developments such as ISO CD/TR 15926 Part 14 [25], there is the concept of a *functional object* and the relation *functionalPartOf* to support this, while the work of [29] applies functions in DOLCE [5] with a role-based approach incorporating elements of social roles [34], conventional-system components [14], and the role-concept, role-player, role-holder triad of [36]. In BFO, there is no such concept as 'functional object' and the BFO concept of 'role' differs such that it is difficult to distinguish where the replaced asset sits in a functional breakdown (as opposed to a physical breakdown). For our purposes, we believe it is sufficient to say that the function of the asset is of the same type for both **a** and **b**, without being specific as to the level of granularity of that type within the functional breakdown (i.e., 'pumping' vs. 'pumping at P101'). Finally, we acknowledge that the idea of replacement is a deeply philosophical problem and raises many questions about identity [11]. However, for the replacement of a limited number of components such as that performed during a replace maintenance activity, this elucidation assumes that the identity of the whole (i.e., the system **s**) remains unchanged.

The semi-formal description for a repair activity describes a process that improves the "ability of an item to perform its required function". For this, and for the maintenance terms that follow, we draw on our previous work on the notion of maintenance state [68]. In this work, we model the states in which items do not have the ability

Table 7

Subject Matter Expert (SME) natural language descriptions, semi-formal descriptions and elucidations for maintenance activity terms

|  |  |
|---|---|
| **Replace** | |
| SME Desc. | Replacement of an item by a new or refurbished item of the same type and make. |
| Semi-formal Desc. | A BFO: Process in which one item is removed and another item with the same required function is installed in its place. |
| Elucidation | **p** is a *replace activity* = <br> Def. **p** is a *process* and there exist *material artifacts* **a** and **b** (**a** $\neq$ **b**) such that **a** and **b** participate in **p** at some time. The following is also true: <br> - **a** has a *function* **f** and **b** has a function **f2**. <br> - **f** and **f2** are of type **F** <br> - **a** is a *continuant part of* a *system* **s** at *time* **t** and **b** is not. <br> - **b** is a *continuant part of* a *system* **s** at *time* **t2** and **a** is not. <br> - **p** *occurs at* some time **tp** <br> - **t** *precedes* **t2**, **tp** *finishes* **t** and **tp** *starts* **t2**. |
| **Repair** | |
| Desc. | Manual maintenance action performed to restore an item to its original appearance or state. |
| Semi-formal Desc. | A BFO:Process, that improves the ability of an item to perform its required function, including restoration of the function. |
| Elucidation | **p** is a *repair activity* = <br> Def **p** is a *process* and there exists a *material artifact* **m** that has a function **f** and participates in **p** at some time. The following is also true: <br> - **m** has state at some time some *failed state* or *degraded state* **s** and *operating state* **s2** <br> - **p** *occurs at* some time **tp** <br> - **s** disables **f**, **tp** finishes **s** and **tp** starts **s2** |
| **Inspect** | |
| SME Desc. | Periodic, careful scrutiny of an item carried out with or without dismantling. |
| Semi-formal Desc. | A BFO:Process, aligned with a preventative maintenance strategy, to estimate the ability of an item to perform its required function. |
| Elucidation | **p** is an *inspect activity* = <br> Def **p** is a *process* and there exists a *material artifact* **m** that has a *function* **f** and participates in **p** at some time. The following is also true: <br> - **m** has state at some time some *state* **s** <br> - **s** is observed by some *observation process* **o** <br> - **p** *is prescribed by* some part of a *preventative maintenance strategy* **ps** |
| **Diagnose** | |
| SME Desc. | The cause of failure, or degradation, is investigated but no maintenance action performed, or action is deferred. |
| Semi-formal Desc. | A BFO:Process, performed on observation of an equipment failure or degradation, to estimate the ability of an item to perform its required function |
| Elucidation | **p** is an *diagnose activity* = <br> Def **p** is a *process* and there exists a *material artifact* **m** that has a *function* (or *capability*) **f** and participates in **p** at some time. The following is also true: <br> - **m** has state at some time some *degraded state* or *failed state state* **s** <br> - **f** *is disabled by* **s** <br> - **s** *is observed in* some *observation process* **o** |

to perform functions (i.e., participate in a process that realizes a specific function). We achieve this through a *disables* object property where some *state disables* a *function realising process*. Furthermore, we determine that in maintenance, when an item is installed, but does not have the ability to perform its required function, it is in a *degraded state* or a *failed state*. Otherwise, the item is in an *operating state* and can perform all of its functions. Given this work, we describe a repair activity as a situation where there is a *material artifact* **m** that participates in

Table 7

(Continued)

### Adjust

| | |
|---|---|
| SME Desc. | Bringing any out of tolerance condition into tolerance. |
| Semi-formal Desc. | A BFO:Process involving a change of state in an item that improves the ability of an item to perform its function or fulfil its capabilities without requiring restoration of the function/capability. |
| Elucidation | **p** is an *adjust activity* =<br>Def **p** is a *process* and there exists a *material artifact* **m** that has a function (or capability) **f** and participates in **p** at some time. The following is also true:<br>- **m** has state at some time some *sub-optimal operating state* **s1** and *optimal operating state* **s2**<br>- **f** *is impacted by* **s1**<br>- **p** *occurs at* some time **tp**<br>- **tp** *finishes* **s1** and **tp** *starts* **s2** |

### Service

| | |
|---|---|
| SME Desc. | Periodic service task: not normally dismantling of the item. |
| Semi-formal Desc. | A BFO:Process, aligned with a preventative maintenance strategy, involving a change of state in an item that involves no change in the ability of the item to perform its function. |
| Elucidation | **p** is a *service activity* =<br>Def **p** is a *process* and there exists a *material artifact* **m** that has a function **f** and participates in **p** at some time. The following is also true:<br>- **m** has state at some time some *operating state* **s** and *operating state* **s2**<br>- **p** *occurs at* some time **tp**: **tp** finishes **s** and **tp** starts **s2**<br>- **p** is prescribed by some part of a *preventative maintenance strategy* **ps** |

### Calibrate

| | |
|---|---|
| SME Desc. | A task involving testing and adjustment of instrumentation to assess and restore function. |
| Semi-formal Desc. | A BFO:Process performed on equipment instrumentation, to estimate the ability of the instrumentation to perform its function and restore if needed. |
| Elucidation | **p** is a *calibrate activity* =<br>Def **p** is a *process* and there exists a *material artifact* **m** that has a function **f** and participates in **p** at some time. The following is also true:<br>- **m** is the bearer of a *control and monitoring role* **r**<br>- **m** has state at some time some *failed state* or *degraded state* or *operating state* **s** and *operating state* **s2**<br>- if **s** is *failed state* or *degraded state* then **s** disables **f**,<br>- if **s** is *sub-optimal operating state* then **s** impacts **f**<br>- **p** *occurs at* some time **tp**: **tp** finishes **s** and **tp** starts **s2**<br>- **p** is prescribed by some part of a *preventative maintenance strategy* **ps** |

this activity and has some function **f**. Before the repair activity, **m** is in some degraded state or failed state **s** (i.e. a state that disables **f**). After the repair activity, **m** is in an *operating state*. Using our modelling of state, we have been able to model changes of state and their implications a result of a replace activity.

An inspect activity is described as a process to *estimate* the ability of an item to perform its required function. Since an item's *state* determines its ability to perform its required function, as described previously, this elucidation describes an *observation process* **o** in which the item's *state* **s** is examined. Note that from a maintenance perspective, an item's ability to perform its required function is not something that is measured (or estimated) on a continuous scale. Instead, the item is either able to perform its function or it is not. Therefore, we believe an observation of the item's state (failed, degraded or operating) is sufficient to model an *inspect activity*.

An inspect activity differs from a diagnose activity in one crucial aspect. In the SME natural language description, an inspect activity is defined as "periodic". This means, that a preventative maintenance strategy is defined that determines at what intervals an item should be inspected. Note that we consider each instantiation of a preventative maintenance activity to be prescribed by the maintenance strategy once it occurs due to its predefined nature. A full definition of maintenance strategy and how it relates to activity types and their individuals is reserved for future work. A diagnosis, in contrast to an inspect, is a *corrective* action to determine "the cause" of a failure (and is not

scheduled). Note that even though an action is *corrective*, it may still be described in a maintenance strategy in the case of a Run-to-Failure corrective maintenance strategy. This distinction between corrective and preventative activities is crucial to our application-level ontology described in Section 5 as, for example, inspect and diagnose are otherwise difficult to distinguish. A key differentiating aspect between inspect and diagnose is when the observation process occurs relative to when the fault is known, which is not something that is generally captured by the data nor an ontology.

An adjust activity is a process that involves changing the state of an item but with no change to the ability of the item to perform its function. This means that adjust activities are performed when an item enters a 'sub-optimal' state, i.e., a state where some aspect has gone out of tolerance but not to a degree that it leads to failure or degradation. The elucidation for an adjust activity looks very similar to a repair activity as they are both corrective actions involving a change of state. The difference is that a *repair activity* requires the item to be in a *failed state* i.e., unable to perform a required function). An adjust activity is performed on items that are still functional but have been observed as being "out of tolerance", as per the SME natural language description in Table 7. The resulting work is corrective. Therefore, we introduce two sub-classes of *operating state* (extending our work in [68]). We model **s1** as a *sub-optimal operating state* (i.e. where equipment is out of tolerance) and **s2** as an *optimal operating state*. As with an inspect activity and a diagnose activity, the difference between an adjust activity and a service activity is that an adjustment is a corrective action, and a service occurs at periodic intervals and is aligned with a preventative maintenance strategy.

The final maintenance activity type is a calibration activity. This is defined as a task involving testing and adjustment of instrumentation to assess and restore its function. In our elucidation, we define a calibration activity similarly to a repair activity, but the item can be in a functioning, degraded or failed state before the activity. What distinguishes a calibration from other activity types is that it is performed on *instrumentation*. Therefore, in the elucidation, we say that the material artifact **m** plays a control and monitoring role **r**. This relationship is important for our use case in Section 5, as we distinguish calibrate activities by their membership in a Control and Motoring sub-unit of the centrifugal pump. Finally, a calibrate activity is preventative as prescribed by some part of a preventative maintenance strategy (as with service and inspect activities).

## 4.3. Reference ontology implementation

The reference ontology is implemented in OWL and is available at https://github.com/uwasystemhealth/Paper_Archive_Maintenance_Activity in the file *maintenance-activity.owl*. A conceptual diagram of this reference ontology is shown in Fig. 2 (where "C" is a class and "@" is an annotation property). In the ontology, there are two classes, Maintenance Activity and Supporting Activity. These activities are a subclass of BFO:Process. The core maintenance activities and supporting activities that were identified in Section 4.1.1 are classes in the ontology and have their SME natural language description, semi-formal description and elucidation stored as annotation properties. The other terms that formed part of each grouping (i.e. align and amend as members of the adjust group) are stored in an annotation property, synonym. "Synonym" is defined in the IOF as "an alternative label (designation)
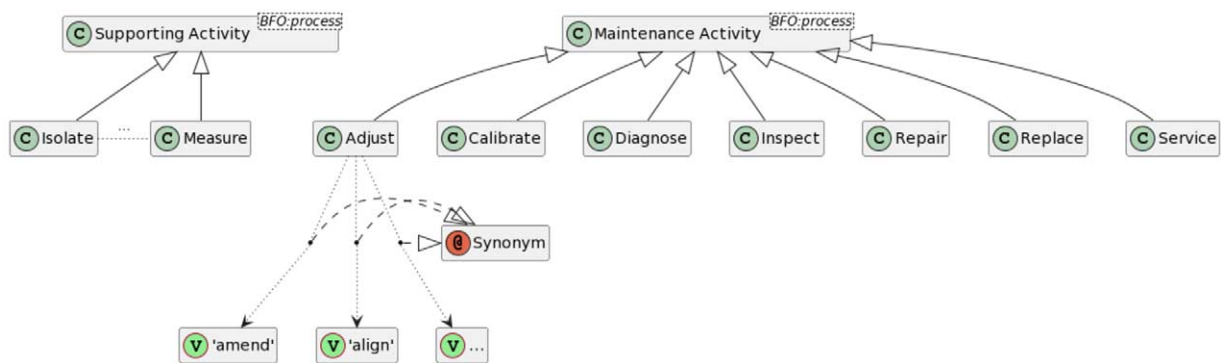


Fig. 2. A conceptual diagram of the maintenance activity reference ontology with example links to synonyms (synonyms are listed in Table 6).

used for the resource in some community" [22]. While it could be argued that the words listed as synonyms are not "true" synonyms with one another in an English language sense, the maintenance community[1] interprets these words in the same way. Therefore, for our maintenance activity reference ontology, this is an appropriate annotation property to use.

## 5. Application-level ontology for MWO data quality

In this section we define an application-level ontology to answer a set of competency questions related to the use case described in Section 5.1. The purpose of this model is to meet a real industrial need while making use of the reference-level terms and ontological analysis described in Section 4. As discussed, the quality of maintenance work order data is an unsolved problem that is imperative for downstream statistical tasks such as mean-time-to-failure analyses. However, maintenance engineers rarely spend time checking the quality of this data, as there is no widely-accepted, repeatable manner to perform this task. This model uses SWRL rules and ontological classification to provide a structured and repeatable way for organisations to assess the quality of maintenance work orders that reflects the logic currently used by engineers performing this task in industry.

### 5.1. Competency questions

The application-level ontology is designed to perform a specific reasoning task to meet a real industrial need. The work order data described in our use case (Section 3), contains seven distinct activity words. These are adjust, calibrate, diagnose, inspect, repair, replace, and service. For data quality purposes, we want to determine if the activity term classified by the ontology is consistent with activity term classified manually by a subject matter expert based on information in the MWO. The competency questions that we use to assess this ontology are:

  Q1. Is the information in a MWO record indicating a replace activity (through various terms) consistent with the expectations of such an activity? Specifically, are the activities associated with MWO records described by 'replace' terms classified by the replace activity type?
  Q2. Is the information in a MWO record indicating a calibrate activity (through various terms) consistent with the expectations of such an activity? Specifically, are the activities associated with MWO records described by 'calibrate' terms classified by the calibrate activity type?
  Q3. Is the information in a MWO record indicating a service activity (through various terms) consistent with the expectations of such an activity? Specifically, are the activities associated with MWO records described by 'service' terms classified by the service activity type?
  Q4. What is the likely activity type for a record in which no discernible activity term was recorded?

### 5.2. Application-level ontology implementation

This ontology was developed making use of the reference ontology described in Section 4. The ontology consists of eight modules, as shown in Fig. 3. A description of each of these modules is contained in Table 8. This ontology can also be found on GitHub at https://github.com/uwasystemhealth/Paper_Archive_Maintenance_Activity. In the following sections, we will discuss the key ontological choices made in the development of this ontology.

We have developed a set of modular, inter-operable ontologies aligned to a top-level ontology (BFO) guided by the processes described in [65]. The diagram in Fig. 3 should be read from top to bottom, first BFO is loaded in Protege, then the IOF Annotation Vocabulary followed by the IOF Core Ontology. The five remaining ontologies are needed for the work in this paper but are reusable in other ontological work. For example, an Asset List, a Pump Functional breakdown description, a list of Maintenance terms and the structure of a maintenance work order are represent generic information that is re-usable. They are currently minimalist ontologies that meet the needs of the use case as the focus of the paper is not the asset classes nor breakdown

---

[1]Possibly the maintenance community within a specific organisation which could tailor the synonym allocations through data profiling on their own MWO records, for example.
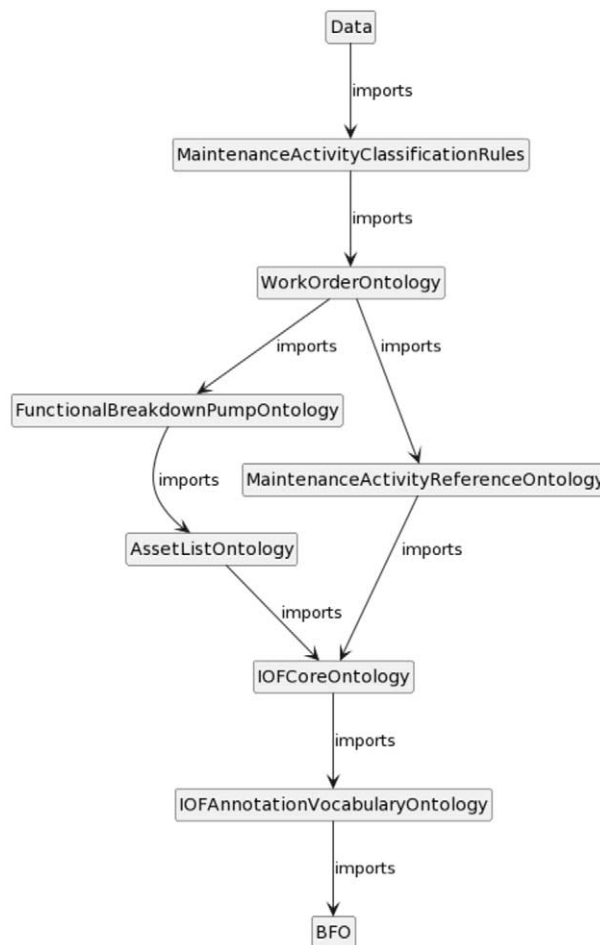
Fig. 3. Import structure of application-level ontology for assessing MWO data quality.

structures. The only ontology specific to the addressing of the use case questions in this paper is the `Maintenance Activity Classification Rules`. Should users want to apply this data quality ontology to another asset class, they potentially only need to extend or replace the `Asset List`, `Pump Functional breakdown`, and `Maintenance Activity Classification Rules` ontologies.

### 5.2.1. Work order records and information content

At the core of this application-level ontology is an Information Content Entity, the MWO Record. In this section, we will explain our choices in modelling a MWO Record and its information parts (i.e. Material Cost, Labor Cost). A conceptual diagram of the information content entities in the application-level ontology can be found in Fig. 4 (where "C" denotes a Class, "I" denotes an Individual and "P" denotes a property). Note that we are choosing not to use the Information Artefact Ontology as a reference ontology, as it does not contribute to the ontology for this use case; however, the classes utilised in our ontology are compatible with it.

As shown in Table 5, a MWO Record has multiple fields (or MWO Data Items that contain information relevant to the MWO. To capture the information contained in these fields, we have added a new object property called refers to that is a sub-property of is about. If an Information Content Entity refers to an entity, then it uses, evokes, or references the entity or concept in some way that is not necessarily descriptive, designative, nor prescriptive.

In this ontology, refers to is used in several cases including:

- Where a Descriptive Information Content Entity, e.g., a MWO Record or MWO Data Item, references an asset in a functional location; the functional location tag field of the record cannot be a designative entity in this case

Table 8

Purpose of each module in the application-level ontology for assessing MWO data quality

| Module | Description |
| --- | --- |
| Data | Contains individuals extracted from MWO data in Table 5. |
| Maintenance Activity Classification Rules | Contains SWRL rules, and additional activity classifiers, used to perform activity classifications according to the workflow in Fig. 1. |
| Work Order Ontology | Contains notions related to MWOs (i.e. Maintenance Work Order Record, Maintenance Work Order Record Data Item, Work Order Execution Process). |
| Functional Breakdown Pump Ontology | Imports the assets contained in the Asset List Ontology and divides them into the subunits shown in Table 4. |
| Asset List Ontology | Contains a list of material artifact classes in a centrifugal pump (e.g. Mechanical Seal). |
| Maintenance Activity Reference Ontology | The reference ontology for maintenance activities described in Section 4. |
| IOF Core Ontology | The IOF Core Reference Ontology retrieved from [22]. |
| IOF Annotation Vocabulary Ontology | The IOF Annotation Vocabulary (includes Synonym, for example) Ontology retrieved from [22]. |
| BFO | The Basic Formal Ontology, used as a foundational ontology [27]. |
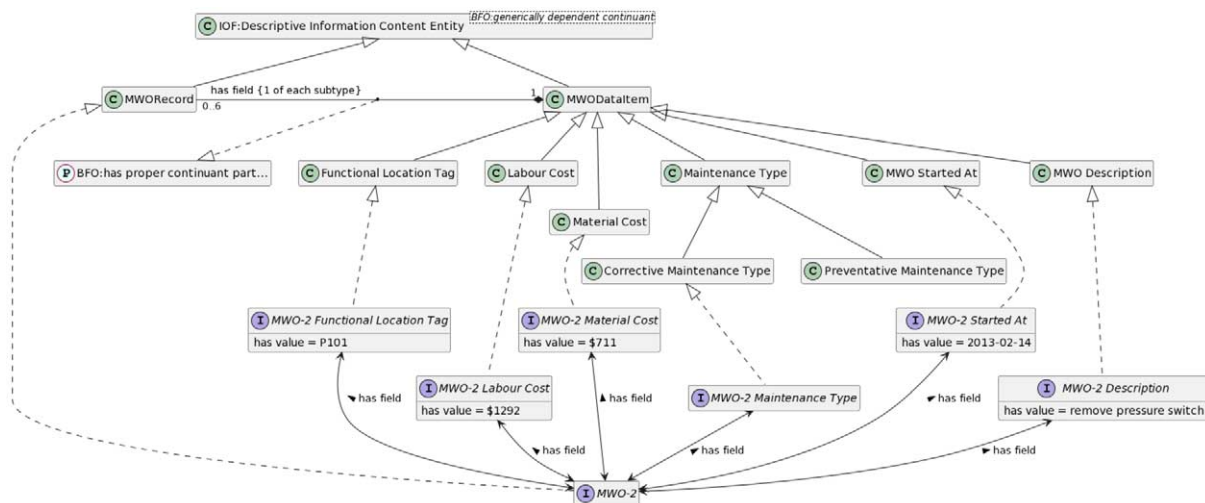


Fig. 4. A conceptual diagram of information content entities for work order records in the application-level ontology.

as the designation of the functional location (and/or the asset installed in the location) comes from elsewhere, the field merely contains a value that matches the designation of the location/asset and, hence, refers to it. This could be a Referential Information Content Entity; however, we do not yet introduce such a class. Moreover, the field cannot be said to describe the functional location/asset as it is describing an aspect of the work order process/activity, by identifying a primary participant in the process/activity.

– Where a Descriptive Information Content Entity such as a MWO Data Item, is deduced to be referencing an entity or concept, e.g., the work order record description field having been identified (through the NLP pre-processing) as referring to a specific *activity* type, which is reconciled from the textual form used in the description and the terms associated with the different activity classes.

Finally, we are choosing not to link the MWO Record (as a generically dependent continuant) entities to their concretizations in the originating database. Instead, we provide only an abstraction/normalisation over the records stored in an actual database. We take for granted, at the moment, that there is an originating database and, hence, some specifically dependent continuant (i.e., the physical datum) and the common independent continuant to which both the datum and the MWO record relate.
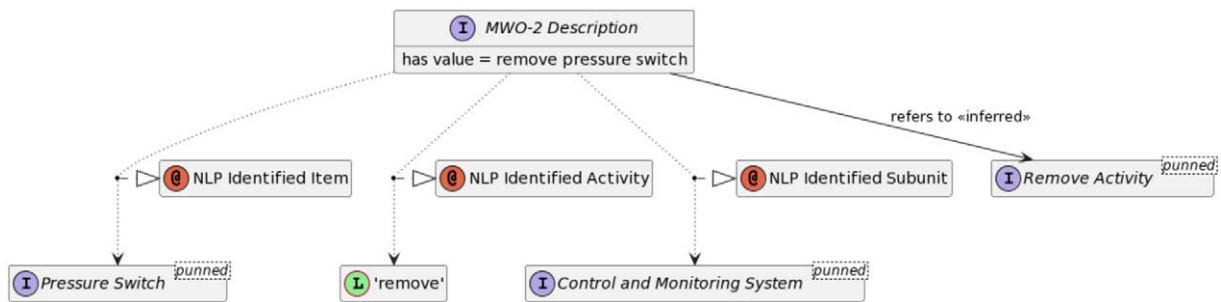
Fig. 5. A conceptual diagram of NLP identified properties for work order descriptions in the application-level ontology.
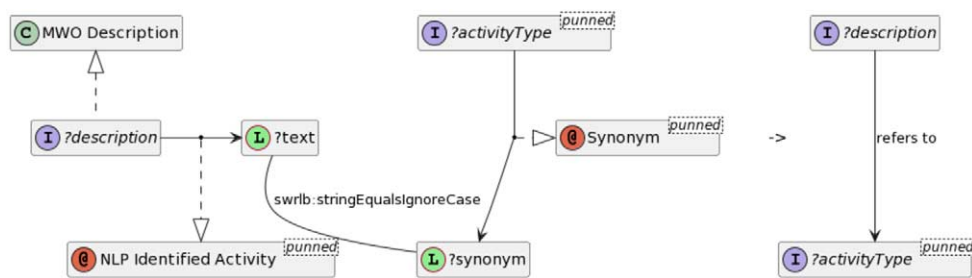


Fig. 6. SWRL rule reconciling NLP extracted activity term with the ontology. Elements representing variables have names beginning with a question mark ('?').

### 5.2.2. Activity terms identified with NLP

As discussed, the goal of this application-level ontology is to check if the activity described in the MWO's unstructured text field matches the information in the rest of the ontology. To achieve this, we store the NLP Identified Activity from Table 5 as an annotation property on the Maintenance Work Order Description. A conceptual diagram of this pattern in shown in Fig. 5 (where "L" represents a 'Literal' property value). We represent this information in annotation properties because they represent *assumed* knowledge resulting from an entity recognition algorithm applied to the specific field. Thus we cannot assert this knowledge as individuals in the ontology without a detailed ontological analysis, which is not necessary for our use case.[2] We use the information stored in this annotation property to infer the NLP identified activity type in the maintenance activity reference ontology described in Section 4. We achieve this inference using punning of the property and activity types, and a SWRL rule that compares the normalised text value of the annotation property against the synonyms of the activity classes. We reconcile term and activity type within the ontology noting that the synonym terms in each group and associated label may be updated in the ontology based on an organisation's profiling of their own data. The SWRL rule is illustrated in Fig. 6.

Using the inferred maintenance activity, we can then compare the NLP identified activity to the result of the workflow in Fig. 1. The rules for which are discussed in Section 5.3. If there is no NLP Identified Activity, the MWO record is considered to refer to an Unspecified Activity. We acknowledge that MWO records with an Unspecified Activity lead to records being considered 'non-matching' in the comparison; however, beneficial analysis can still result by flagging such records for review by an engineer. In the following section, we will describe how the NLP Identified Item and the NLP Identified subunit are used to gain further information about the *true* activity performed in the work order.

---

[2]Note that the OWL implementation provided does not use annotation properties due to errors occurring in the use of the tools. Repeated savings of the file would cause the annotation properties to shift into data or object properties in ways that would break the outcomes of the reasoning. So, the authors defined them as data/object properties to avoid such recurring issues. We assume that different tools would not demonstrate the same issues and the properties could be defined as annotation properties as described here.
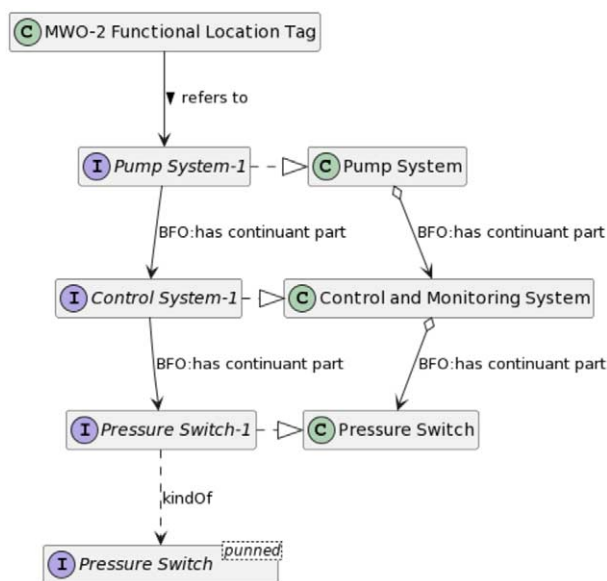
Fig. 7. A conceptual diagram of items and subunits in the application-level ontology.

### 5.2.3. Items and subunits

To realise the logic contained in the workflow in Fig. 1, we need to know the type of equipment being worked on (or, more specifically, which subunit from Table 4 the item fits into). In the MWO structured data, there is a functional location field (a Functional Location Tag in our ontology). For all of the work orders in this use case, the Functional Location Tag refers to a Pump System (a centrifugal pump and other equipment e.g, motor, values, instrumentation necessary to perform the pumping function). The Pump System and other item classes are defined as IOF Core Material Artifacts, which is a BFO Object class. In our ontology, a Pump System has its (functional) subunits (e.g. the Control and Monitoring System) as continuant parts *at all times*. The subunits, modelled as Engineered System classes under the BFO obect aggregate hierarchy, have continuant parts *at some time*, e.g., the Control and Monitoring System having further continuant parts such as a Pressure Switch and other types of instrumentation. This is a minimal representation of the asset's functional breakdown structure, which we do not address further as it is not the focus of this paper and is a complex topic to address in BFO. A conceptual view of items and subunits in the application-level ontology is shown in Fig. 7. For reasoning purposes, the item and subunit classes are punned to individuals and matching values for the object property *kindOf* are inferred by the classification. The equipment items are parts, the classes for which are contained in the Asset List Ontology module while the equipment's classes membership in a subunit is captured in the Functional Breakdown Ontology (as described in Table 8) – the punning and *kindOf* property are separated into the Maintenance Activity Classification Rules module.

The only structured information in the MWO record that indicates the item being acted upon is the Functional Location Tag. Since this is not at the right level of granularity to perform our analysis, we rely on our entity recognition algorithm, and some subsequent analysis, to identify the specific item and subunit. For the purpose of this analysis, we use a simple mapping from the item type to subunit type. The breakdown structure of centrifugal pumps identified in Table 4 has a unique subunit type for each maintainable item type. Using the NLP Identified Item (model shown in Fig. 5), we can utilise the subunit information in the rules to identify possible participants of the activity, i.e., the item individuals at the part level, and further use this to infer a classification for the maintenance activity that was performed on the item.

### 5.3. Maintenance work order execution processes and activities

Thus far, we have shown how we model the structured information from the work orders as Information Content Entities. Subsequently, we have shown how we infer the *expected* maintenance activity that was performed on the
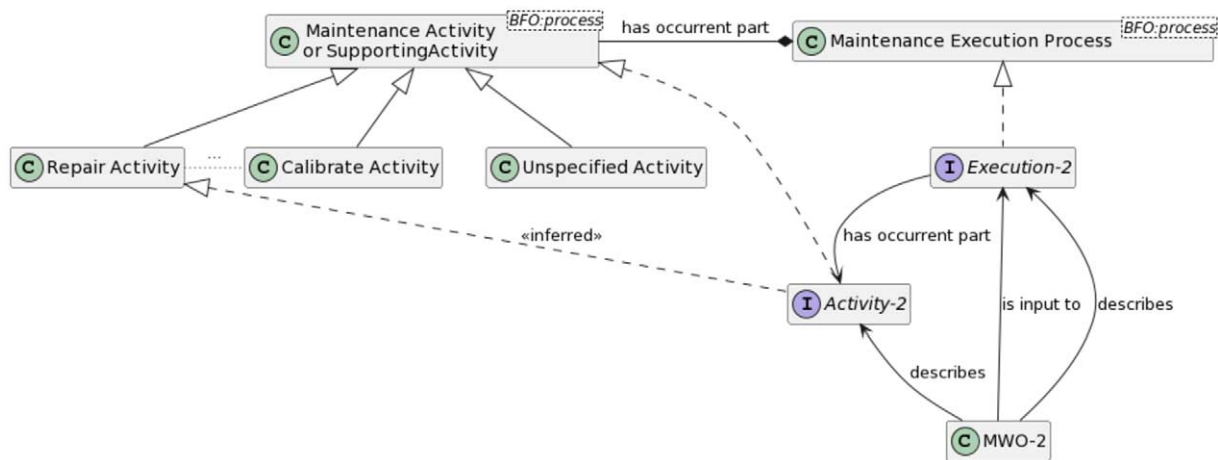
Fig. 8. A conceptual diagram of maintenance work order execution processes and activities in the application-level ontology.

item (from the NLP Identified Activity). Finally, we have shown how we use extra data from our unstructured text by storing the item being acted on and the item's subunit. In this section, we show how we use this information to realise the logic shown in Fig. 13 and infer the type of maintenance activity that was likely to have been performed. If the type of maintenance activity identified through analysis of the description does not match the inferred maintenance activity classification, we assume that the MWO data is not of a high quality and requires further human attention.

The conceptual diagram in Fig. 8 shows how we model the inference of the activity type and its relation to a MWO record. In the figure, our Maintenance Work Order Record (MWO-2) describes some Maintenance Work Order Execution Process, following the structure of the ROMAIN ontology [30]. This execution process represents all the activities that are carried out to execute a MWO (and is contained in the Maintenance Work Order Ontology from Table 8). This execution process has the Maintenance Activity to be classified as an occurrent part. In the case of an activity that was not actioned, the classification is applied to the execution process and not the activity. This is because the activity cannot really exist if the activity to be classified was not actioned, but other parts of the process may have been. This way the individual for the unactioned activity could be removed if desired.

The actual maintenance activity classification is inferred using SWRL rules that capture the logic of the workflow shown in Fig. 13. To support this inference, there are several supporting elements, including additional classifiers and rules, that surface information used by other rules or for subsequent querying. The primary rules and supporting elements are captured in the Maintenance Activity Classification Rules module, described in Table 8.

### 5.3.1. Additional activity classifiers

There are two main categories of additional activity classifiers: intermediate classifiers, and final classifiers. The latter correspond to classifications inferred at the leaf nodes of the workflow (refer Fig. 1), while the former may be inferred at non-leaf nodes. Furthermore, some of the final classifiers are intended for querying purposes (namely those of the Inferred Activity category) while others represent the outcome of the decision logic.

The interim classifiers are based on the subsets of activity class they encompass and include:

– Maintenance Supporting or Unspecified Activity (Maintenance Activity, Supporting Activity, or Unspecified Activity)
– Corrective Action (adjust,diagnose, repair, or replace)
– Corrective Action with Low Material Cost (adjust, or diagnose)
– Preventative Action (calibrate, inspect, replace, or service)
– Preventative Action with Low Material Cost (calibrate, inspect, or service)

While the Corrective Action and Preventative Action classifications in the rules could have been inferred through an OWL *equivalent class* axioms, it was chosen to use rules for consistency with the other rules that could not do so.

The final classifiers include a pair of disjunctive classifiers, to meet the decision logic as best as can be where there is not enough information from the MWO records and explicit knowledge from the ontology to make a complete classification. In addition, there are marker classes to indicate the status (i.e., match, non-match, uncertain) of the classification result. These include:

– Adjust or Check – disjunctive class indicating that the activity may be either an adjust activity or a diagnose activity, but the currently available information and decision logic cannot ascertain for certain.
– Repair or Replace – disjunctive class indicating that the activity may be either a repair activity or a replace activity, but the currently available information and decision logic cannot ascertain for certain.
– Inferred Activity – classifier marking activities that have had their classification inferred; is separated into several subclasses:

  ∗ Inferred Type Matches Work Order Description – classifier marking activities as matching the NLP Identified Activity of the MWO record's description field.
  ∗ Inferred Type Does Not Match Work Order Description – classifier for marking activities as not matching the MWO's description field; due to the restrictions of OWL DL and SWRL, this is not used by the rules but can be used by subsequent queries and assertions to track such activities more readily.
  ∗ Uncertain Maintenance Activity – classifier marking activities that are not clearly distinguished as a specific activity type, for example, it will accompany Adjust or diagnose and Unspecified Activity classifications.

### 5.3.2. Additional equipment classifiers

For ease of writing the SWRL rules, some additional classifiers are used to identify particular subsets of equipment subunits required by the rules; these are identified through the process of capturing the expert knowledge used by the classification rules. The classes are quite general and apply orthogonal classifications without restricting the primary classification of subunit. As such, they should be broadly applicable to categories of equipment or, at the very least, the classes of pump other than centrifugal pump. This should control the number of classes required when defining activity classification rules for broader classes of equipment than illustrated here. For example, it does not matter if a particular class of pump does not include a subunit of a particular type; only if there is a contradiction in the types of subunit relevant to particular actions would more fine-grained distinctions need to be made. While it is possible to directly identify the class expressions in the rules, providing named classes is more flexible, convenient, and manageable. The classes include:

– Pump Subunit – disjunction of Engineered System subclasses that are can be included in Pumps as subunits.

  ∗ Inspectable Unit – subunit classifier identifier subunits that are relevant to inspection activities, i.e., not Lubrication System, nor Control and Monitoring System, nor Driver and Electrical System
  ∗ Not Control And Monitoring System – some rules make reference to any subunit other than a Control and Monitoring System: this class captures that requirement
  ∗ Not Pump Unit System – some rules make reference to any subunit other than the Pump Unit System: this class captures that requirement

### 5.3.3. Item material and service costs

The decision logic for classifying the activities relies quite heavily on the material costs of the work that was performed. These costs are proxies used to evaluate whether work was performed and whether it was of the expected type. According to the workflow of Fig. 13, there are two different material costs that are evaluated $X$ and $Y$, called here *material item cost* and *item service cost*, respectively. The material item cost is an indicator of the expected minimum cost if an item were to be replaced/repaired, while the service cost is an indicator of the minimum costs associated with servicing an item. A pair of data properties is used to capture this information on a per item basis (likely derived from the item's class). Such an approach makes the rules more flexible as specific values are not hard coded into them.

Of course, the item cost and service cost information is not currently in the scope of the available data. However, this information could be derived from other sources, data profiling, or other means of estimating the costs associated with activities related to equipment. If available, such information can be incorporated without modifying the rules to improve the granularity of the activity classification rules. Since we do not currently have the detailed information,

we assume that the current data is indicative of the maintenance activities performed on the centrifugal pumps of this organisation and set default values accordingly: material item cost is $150, while item service cost is set at $1.

The possibility of avoiding the material costs, as it pertains to replacements, is discussed in Section 7.2 as it is unlikely to be the most accurate approach in general.

### 5.3.4. SWRL rules

In total there are 13 rules for classifying activities according to the leaf nodes of the workflow (Fig. 13), 4 additional rules inferring intermediate classifications, and 2 rules for final comparison of classification to the NLP Identified Activity type. Figure 9 illustrates a simple rule for identifying the corrective maintenance branch of the workflow.

Intermediate classifications cannot be fully relied upon due to some of the activity classes falling under both corrective and preventative maintenance categories. This application ontology does not address the ontological concern of the activity individual being a corrective or preventative occurrence, it merely uses the classifications as indicators. Therefore, the actual maintenance type of the associated MWO record is relied upon to ensure that the final classifications are on the correct path of the tree where activity classes that may be of either category are involved. Capturing it in a rule also makes it easier to reuse the pattern in other, more complex, rules. Figure 10 illustrates a subsequent terminal rule that identifies the activity as repair or replace, alongside a marker classification indicating that it is uncertain as it cannot fully distinguish between the repair and replace activity types.

The rule for the final cross-check for matching the activity classification is shown in Fig. 11.
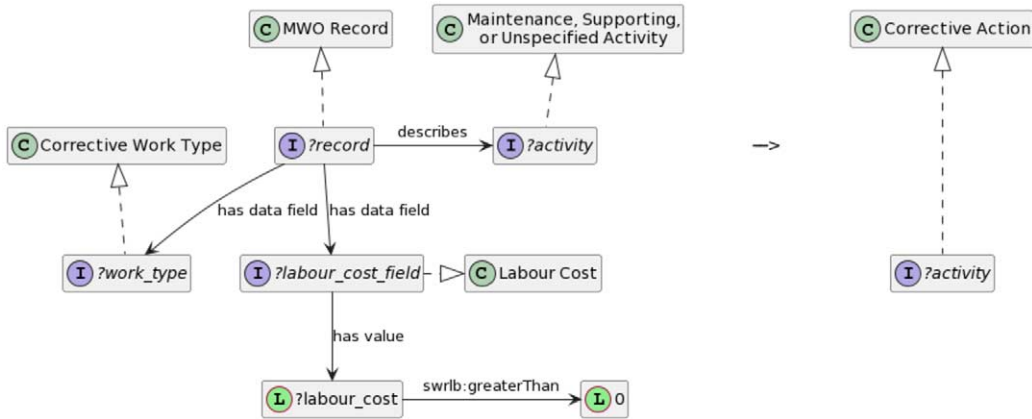


Fig. 9. Rule for inferring the Corrective Action classification based on the information in the MWO record.
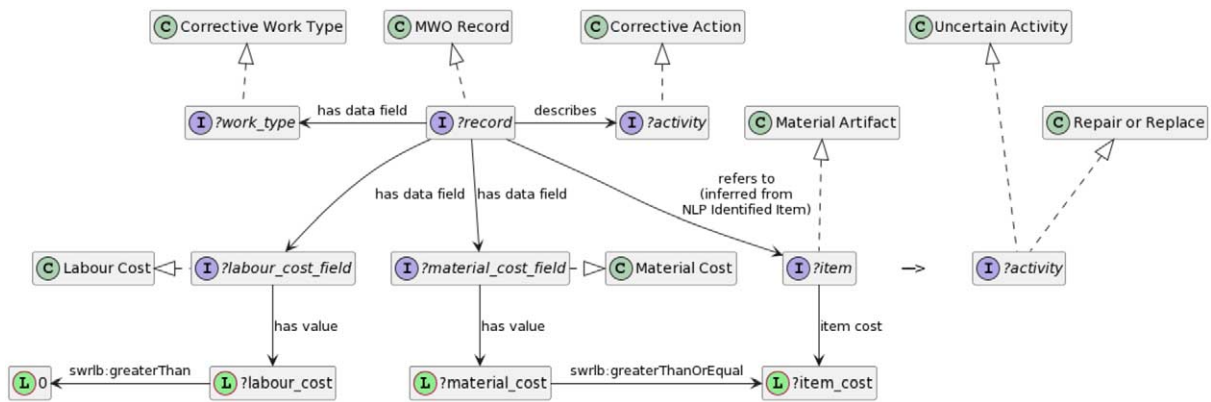


Fig. 10. Rule inferring the Repair or Replace classification, including the marker classification Uncertain Activity, based on the information in the MWO record.
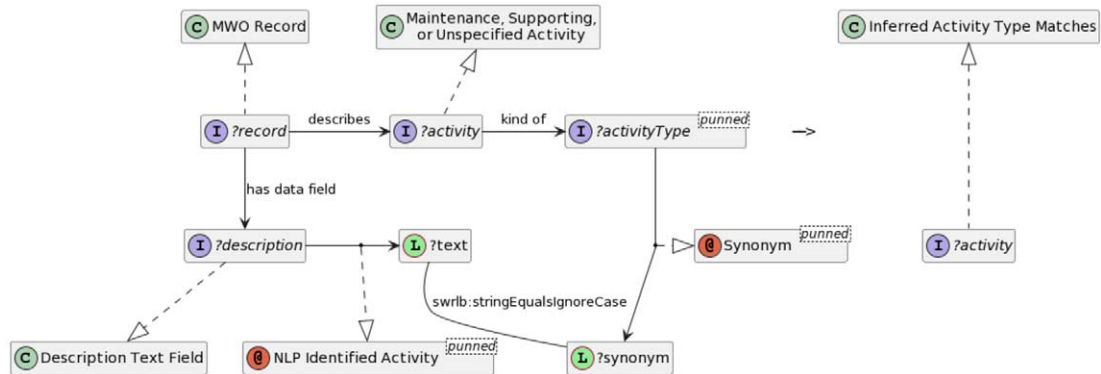
Fig. 11. Rule inferring the match between a concrete activity classification and the NLP Identified Activity type.

```
PREFIX core:<http://www.industrialontologies.org/core/>
PREFIX activity:<http://www.semanticweb.org/maintenance-activity#>
PREFIX rule:<http://www.semanticweb.org/maintenance-activity-classification-rules#>
SELECT DISTINCT ?activity ?clazz ?nlpActivityType
WHERE {
  ?activity a/rdfs:subClassOf* rule:InferredActivity ;
    a ?clazz .
  FILTER ( ?clazz != rule:InferredActivity &&
           ?clazz != rule:UncertainActivity &&
           ?clazz != rule:MaintenanceSupportingOrUnspecifiedActivity )
  FILTER NOT EXISTS {
    ?activity a/rdfs:subClassOf* rule:InferredActivityTypeMatchesWorkOrderDescription
  }
  OPTIONAL {
    ?activity core:describedBy|^core:describes ?_record .
    ?_record rule:refers_to ?nlpActivityType .
    ?nlpActivityType rdfs:subClassOf+ activity:MaintenanceActivity .
  }
}
```

Fig. 12. SPARQL Query for retrieving the activities that did *not* match the NLP Identified Activity type.

*Querying the classifications with SPARQL*    Using the inferred information, the inferred activity classifications can be queried including, but not limited to, all inferred activity classifications, activity classifications matching the NLP Identified Activity information, uncertain activity classifications, and non-matching activity classifications. Of particular interest are the non-matching, including uncertain, activity classifications. Due to the constraints of OWL DL and SWRL, retrieving the non-matching activity classifications leverages the Inferred Type Matches Work Order Description class and SPARQL negation to match the non-existence of the pattern in which it occurs. An example query is shown in Fig. 12.

The query identifies all the activity individuals that have an inferred activity classification, a Maintenance Activity class, that does not match that class indicated by the NLP Identified Activity. As output, it includes the classifications of the activity individual as well as the activity class associated with the MWO record (via the NLP Identified Activity property). The latter part of the query is optional so as to include results where the activity class associated with the MWO record is missing, in case there is an error in the process. Such a query allows an engineer to review the non-matching, including uncertain, classifications against what was written in the MWO record; they can then take action to rectify errors to improve the data quality.

## 6. Evaluation

We verify this ontology using the OOPS! Ontology Pitfall Scanner [45]. Issues identified by the scanner with a severity of "important" and "critical" have been considered and addressed where appropriate for this ontology.

We validate this ontology by 1) comparing the results of classification of the activity term in each work order by the ontology and subject matter expert (the agreement is either yes, no or partial), and 2) using a confusion matrix to examine the performance of the ontology and subject matter expert in identifying a data quality issue. A data quality issue is identified when the activity term in the original text of the work order disagrees with the result of ontology and/or subject matter expert's reasoning. The results of the classification activity are shown in Table 9 and the confusion matrix is shown in Table 10.

In order to assess performance a ground truth for the activity term needs to be established. This was done through consultation with reliability engineers whose job it is to manually review these work orders and extract activity terms to support their analysis as discussed in Section 1, This reasoning process is described below.

When reviewing maintenance work order records, these subject matter experts use the data in the work order fields, their intuition and other explicit information (e.g., parts and labour costs) to determine what activity was actually performed. An example of the logic used by engineers to assess if the description provided by the author of the maintenance work order is consistent with data in other fields of the work order is shown in Fig. 13. First, the item that work was performed on is identified. This might be at the equipment level, e.g., pump, or at the maintainable item level, e.g., pressure switch. Items at the maintainable item level can be located in their respective sub-classes using standardised asset class hierarchies such as that shown in Table 4.

Next a check is made to see if the labour cost field has a value greater than zero dollars. This is necessary to determine if the proposed activity occurred or not. While reliability engineers are interested in both cases, our interest from a data quality perspective is on MWOs that are actioned. The next decision point is determining if the *work type* is preventative or corrective. This allows for a split into two groups. Repair, adjust and diagnose are activities arising only from corrective strategies, Service, calibrate, and inspect are part of preventative strategies, while replace can be either corrective or preventative, the latter as part of an interval-based replacement strategy. The presence and magnitude of parts and labour costs are often used to make inferences about corrective work. For example, a replace activity will have a parts cost equal to the cost of the unit whereas a repair parts cost will be a lot less. This is shown in Fig. 1 by using material cost variables such as $X$ and $Y$ where $X$ would indicate substantial cost associated with a replacement of a unit whereas $Y$ would be associated with use of consumables (filters, belts, etc.) or maintainable items. An experienced engineer will have this sort of information. Diagnose and adjust will incur no parts cost and in this use case the information necessary to separate between these two activities is not available. An adjust activity will have a check element preceding the adjust step to correct an out-of-tolerance condition. Diagnose is an observation process and does not involve the change of state of the item. In this use case we do not have sufficient information to distinguish between adjust and diagnose. Distinguishing between the classes of service, inspect and calibrate is done, in this use case, based on the item and to which sub-unit it belongs; for example, calibrate is associated with items in the Control & Monitoring subclass. The reasoning here is representative of steps taken by reliability engineers every day as they manually process MWOs [15]. Each competency question is addressed in turn in the section below.

### 6.1. Classification results for each maintenance work order

Table 9 is used to identify data quality issues in each record. The table shows for each record a) the activity term in the original MWO unstructured text, b) the classification by a subject matter expert (SME), and c) the activity classification inferred by the ontology and rules. The 5th column shows the agreement between the SME and the ontology as Yes (Y), No (N) and Partial (P). The Partial is used to capture the situation when the ontology identifies two possible activities, one of which was identified by the SME, for example repair or replace and adjust or diagnose. The 6th column captures whether a data quality issue has been identified by comparing the SME classification to the activity term in the original text.
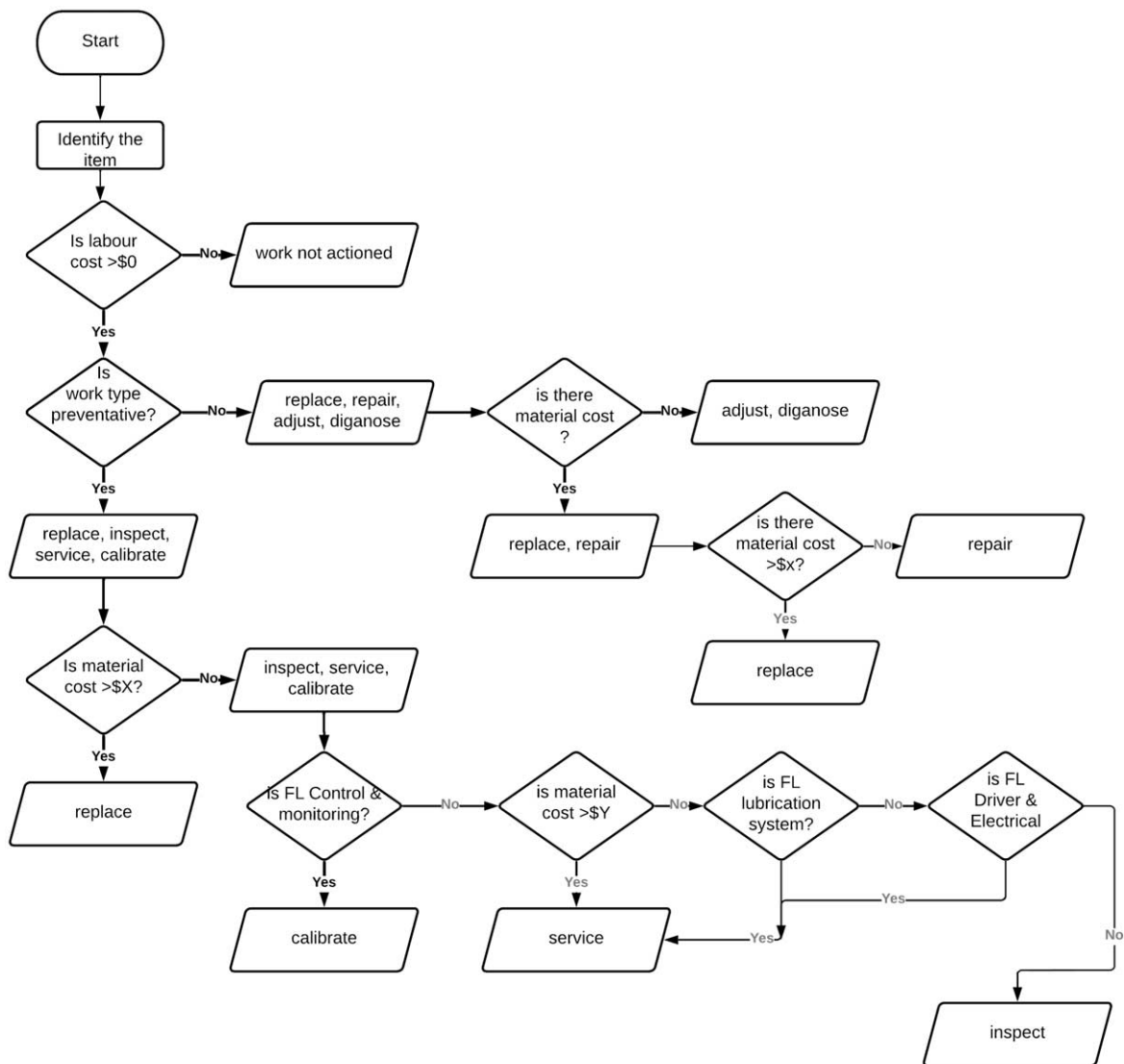
Fig. 13. Classification workflow capturing the logic used by a reliability engineer when interpreting text in maintenance work order records from commonly available information.

### 6.2. Confusion matrix results

A summary of the data quality issues identified by the ontology is shown in a Confusion Matrix in Table 10. For the 36 records in this pump MWO data set from 2012–2020, 20 records (55%) were identified by both the ontology and the expert as having data quality (DQ) issues. There were 14 records with no DQ issues and 2 records in which the ontology and the expert disagree. An analysis of the identification of maintenance activity class by the ontology compared to the class described in the original (NLP-processed) record is shown in Table 11.

### 6.3. How did the ontology perform for on the competency questions?

**Question 1.** Is the information in a MWO record indicating a replace activity (through various terms) consistent with the expectations of such an activity? Specifically, are the activities associated with MWO records described by 'replace' terms classified by the replace activity type?

Table 9

Results of activity classification by the ontology and rules comparing the ontology outputs with the SME classification showing agreement as yes (Y), no (N) and partial (P). The table also includes outputs from the NLP on the original record including activity, item and corrective and preventative (C, P) status

| ID | Activity term in original text | SME classification | Ontology classification | Agreement SME-ontology | DQ issue (original) |
|---|---|---|---|---|---|
| 1 | none (switch)(C) | replace | repair or replace | P | Y |
| 2 | remove (switch)(C) | replace | repair or replace | P | Y |
| 3 | calibrate (switch)(C) | adjust | adjust or diagnose | P | Y |
| 4 | none (pump)(C) | | adjust or diagnose | P | Y |
| 5 | none (switch)(C) | replace | repair or replace | P | Y |
| 6 | calibrate (switch)(C) | work not actioned | work not actioned | Y | Y |
| 7 | service (motor)(P) | service | service | Y | N |
| 8 | calibrate (switch)(C) | adjust | adjust or diagnose | P | Y |
| 9 | calibrate (switch)(C) | adjust | adjust or diagnose | P | Y |
| 10 | calibrate (switch)(C) | replace | repair or replace | P | Y |
| 11 | replace (valve)(C) | replace | repair or replace | P | N |
| 12 | investigate (oil)(C) | work not actioned | work not actioned | Y | Y |
| 13 | none (mech seal)(C) | replace | repair or replace | P | Y |
| 14 | none (switch)(C) | replace | repair or replace | P | Y |
| 15 | check (switch)(C) | diagnose | adjust or diagnose | P | Y |
| 16 | none (pump)(C) | diagnose | adjust or diagnose | P | Y |
| 17 | install (switch)(C) | replace | adjust or diagnose | N | Y |
| 18 | service (motor)(P) | service | service | Y | N |
| 19 | service (pump)(P) | work not actioned | work not actioned | Y | Y |
| 20 | none (switch)(C) | diagnose | adjust or diagnose | P | Y |
| 21 | replace (switch)(C) | replace | repair or replace | P | N |
| 22 | none (motor)(C) | diagnose | adjust or diagnose | P | Y |
| 23 | repair (switch)(C) | repair | adjust or diagnose | N | Y |
| 24 | service (pump)(P) | service | service | Y | N |
| 25 | service (pump)(P) | service | service | Y | N |
| 26 | service (motor)(P) | service | service | Y | N |
| 27 | replace (pump)(C) | replace | repair or replace | P | Y |
| 28 | service (motor)(P) | service | service | Y | N |
| 29 | change (oil)(C) | adjust | adjust or diagnose | P | Y |
| 30 | none (mech seal)(C) | replace | replace | Y | Y |
| 31 | none (pump)(C) | diagnose | adjust or diagnose | P | Y |
| 32 | service (motor)(P) | service | service | Y | N |
| 33 | none (flowmeter)(C) | diagnose | adjust or diagnose | P | Y |
| 34 | repair (pump)(C) | repair | repair or replace | P | N |
| 35 | none (switch)(C) | diagnose | adjust or diagnose | P | Y |
| 36 | replace (switch)(C) | replace | repair or replace | P | N |

Table 10

Confusion matrix: true positive – the expert and the ontology both identify a data quality (DQ) issue. True negative – the expert and the ontology both identify there is no DQ issue. False positive – the ontology identifies a DQ issue but the expert does not. False negative – the expert identifies a DQ issue but the ontology does not

| | Ontology identifies DQ issue | Ontology says no DQ issue |
|---|---|---|
| Expert identifies DQ issue | 20 | 0 |
| Expert says no DQ issue | 2 | 14 |

Table 11

A comparison of the text in the original record (rows) with the classification provided by the ontology (columns) for each activity class

|  | Repair or replace | Adjust or diagnose | Calibrate | Service | Inspect | No activity |
|---|---|---|---|---|---|---|
| Repair | 1 | 1 |  |  |  |  |
| Replace | 5 | 1 |  |  |  |  |
| Adjust |  | 1 |  |  |  |  |
| Calibrate | 1 | 3 |  |  |  | 1 |
| Diagnose |  | 1 |  |  |  | 1 |
| Inspect |  |  |  |  |  |  |
| Service |  |  |  | 7 |  | 1 |
| No activity | 6 | 6 |  |  |  |  |

All original records with terms repair, replace or remove were identified as repair or replace by the ontology and expert. One record (17) uses the verb install in the original record, this leads the expert to classify this as a replace activity whereas the ontology classifies this an 'adjust or diagnose'. The reasoning from the expert is that the work order could be describing a situation in which the old unit was removed (based on an earlier work order) and this work order is to install a new one as part of a replacement activity. There is no material cost associated with this. This could be because the material cost was charged on the original work order associated with the removal task. This could be resolved with access to more information not available here. The ontology flags this as a data quality issue but the expert does not.

**Question 2.** Is the information in a MWO record indicating a calibrate activity (through various terms) consistent with the expectations of such an activity? Specifically, are the activities associated with MWO records described by 'calibrate' terms classified by the calibrate activity type?

This is discussed in Section 7.1.

**Question 3.** Is the information in a MWO record indicating a service activity (through various terms) consistent with the expectations of such an activity? Specifically, are the activities associated with MWO records described by "service" terms classified by the service activity type?

There is high agreement between the expert and ontology with the original text on the activity service with the exception of one record (19) when no work was actioned.

**Question 4.** What is the likely activity type for a record in which no discernible activity term was recorded?

This is discussed in Section 7.1.

## 7. Discussion

### 7.1. What worked well?

**Identifying texts with no data quality issues**: There are 16 texts that the expert identifies as having no data quality issues. In 14 of these cases (88%) the ontology agrees.

**Assigning an activity when no activity is identified in the original text**: One of the main issues in the original text is the absence of a verb describing the desired activity. This occurs in 12 records (33%) in this corpus. This is a common issue in maintenance work orders given there are two reasons for work order generation. As mentioned earlier the first reason is an instruction to perform work to 'change (or restore) of state' for the item and the second reason is a discovery process to determine if work is required. It is the latter that often lacks a verb as shown, for example, in record 1 "pressure switch undersize". There is considerable value in being able to infer what maintenance activity is likely to have been actioned to correct this situation. In this the ontology performs well with 100% agreement with the expert.

**Reclassification of calibration**: The action of calibration involves the "testing and adjustment of instrumentation to assess and restore function" as per Table 7. Calibration is an action prescribed as part of a preventative maintenance strategy. However, on 4 occasions in this data set the work order required corrective work on instrumentation to adjust and restore function. The technician use the word "calibrate" in the original text but the expert and ontology recognise that this is incorrect as the work done is corrective. There is business value in being able to identify these situations using AI, as they indicate, in this case that there should be a preventative maintenance task scheduled for calibration but since there is not, the necessary action is being done in a reactive rather than proactive fashion.

*7.2. What could be improved?*

**Distinguishing between repair and replace for corrective actions**: For many of the work order classifications, there is insufficient evidence to distinguish between two of the activities (i.e. Repair or Replace and Adjust or diagnose) without more asset and site specific information. It would be possible to include additional reasoning between repair and replace based on the material cost. If the material cost is equal to the cost of a new unit then the activity would be replace, if not, then it would be a repair. Other options would involve considering the parts assigned to the maintenance work order to see if these parts included, say, a pressure switch (as a replaced component) or not. This would require adding new classes capturing the parts used and resources consumed in the activity. If the parts included a pressure switch, for example, then we could reason that a replace activity had occurred. If not, then the activity would be considered a repair. The parts list associated for each MWO number is captured in a separate table in the database. While these improvements could potentially produce a more generalised approach, this data was not available for the construction of our application-level ontology.

**Disambiguation of equipment subunits for complex equipment examples (i.e. complex pumps)**: In this paper we have worked with a use case on a centrifugal pump. In the process industries there are a number of equipment classes that have well defined taxonomies of functional subunits and maintainable items. For example, heat exchangers, compressors, pumps, and turbines. CMMS systems in different organisations are set up to use these hierarchies when they have to share data for reference data sets [23]. While the equipment subdivision in Table 4 is fit-for-purpose for standard centrifugal pumps, some specialist pumps might also include a separate pressurized lube oil pump system. Records of maintenance work on these can be difficult for an engineer to disambiguate as there are two pumps in the system and a number of seals, some on the main pump and others on the lube pump. Our focus in this work is on the data quality of the standardised equipment classes which make up the majority of assets in process plants and not on the specialised (and often may be expensive one-of-a-kind) assets. For the latter we recommend that specialist engineers continue to manually review the MWOs. Future work may also investigate means of flagging the ambiguity for review by engineers, potentially with a suggestion of the most likely candidate.

**Disambiguation of Semantic Synonyms for Activities:** In this work we have mapped each of the semantic synonyms to one activity class, as shown in Table 6. While this allows us to move forward with supporting data quality work right now we recognise there could be additional information incorporated into this mapping activity and this will be the subject of future collaborative work in both the NLP and ontology communities. For example, we have mapped install to the replace activity group. We do not know if it is a like-for-like replacement from the information we have in this case study. We have accounted for future work in this area in the elucidation for the term replace shown in Table 7. This would allow reasoning to be developed for information about the parts that were replaced and the parts installed, when this information is available.

**Handling of temporal aspects of activities**: For the current use case we have elided complex reasoning on the temporal nature of the activities. The approach is successful in applying limited information from MWO records to classify the activities; however, it leaves elements, such as the actual participants of the activities, as representative (or possible) participants, as opposed to actually linking the activities to the parts and items involved in the activities. Furthermore, the start and end date/times are not currently attached to the activity individuals, only the MWO records. As additional information is made available and integrated, further reasoning could be performed to tie the two together. This is another avenue of future work.

**Technical Limitations and Assumptions**: the evaluation of the ontology was performed using commonly adopted Semantic Web tools for research and development. These tools exhibit a less than desirable execution time for reasoning over the ontology – approximately one minute per MWO record. We assume that this is due to

technical limitations in their implementations and that production grade Semantic Web databases and tooling would perform much better. If not, other alternatives could be implemented following the same conceptual pattern as illustrated in this paper: e.g., using SPARQL CONSTRUCT queries to realise inferred relations, or translate into a logic programming language for rule execution and reasoning if necessary.

### 7.3. What insights did we get into maintenance activity classification

One insight gained in this work is the value of determining if maintenance activities are corrective or preventative. This information is contained in the *work order type* field in Table 5. When writing elucidations for the terms in our reference ontology, we found that SME natural language descriptions for some activities (derived from descriptions given in international standards), diverge only through the word "periodic". For example, a service activity is a periodic adjust activity. This means that service activities are aligned to a preventative maintenance strategy, and adjust activities are not. This insight informed a key part of the reasoning in the application-level ontology. This insight is consistent with the need for reliability engineers to know if preventative work is being done or not as a check of compliance with maintenance strategy.

Another insight is the importance of an item's state in distinguishing between activity types. To realise this, we link our elucidations to our previous work on Maintenance State [68]. For example, we determined that before a replace activity, an item is in a Degraded State or a Failed State, meaning that the item's "ability to perform its required function" (as per the semi-formal description) is impaired. After the replace activity, the item is in a Operating State as its required function has been "restored". However, the state of an item is not changed throughout an adjust activity. We have used this to distinguish between a repair and adjust activity from an ontological perspective. However, in practice, there is a grey zone between what some would call an adjustment and others a repair. This is worth considering in future work.

In our use case and this application-level ontology, we do not have an item's state information available to us (as condition and performance data is captured in a different data base, not generally in a MWO record). However with the increased emphasis on condition monitoring on assets and process instrumentation of systems this information could be available but would need data from data bases beyond the CMMS. Future work should look at how to use this process and condition data to assist with the decisions being made in this application-level ontology. Digital twins will have to be able to incorporate changes in state resulting from maintenance activities in their models.

### 7.4. Choice of upper ontology

In Section 4.2 we discussed the limitation we encountered with using BFO for the notion of *functional object* and its associated hierarchies. The concept of an asset having both a functional and a physical hierarchy is central to engineering design and systems engineering [4]. In early design, operation and reliability engineers pay significant attention to the functional perspective, identifying what desired and secondary functions are required and how these functions are fulfilled. Functional units, in engineering, are usually associated with desired levels of performance. As the design life cycle proceeds towards manufacturing then physical and structural consideration get increasing attention. Design for maintenance also considers the physical structure of the system, subunits and maintainable items. For ontologies to be applied to industry data, the models and reasoning must be able to accommodate both functional and physical perspectives. The development of an ontological understanding of functional hierarchy is still an ongoing research line with relevant work available from [4,5].

The authors of this paper are part of IOF Maintenance working group and this relationship was the primary reason for selecting BFO as the upper ontology. Our interest is in understanding performance of this upper ontology on a range of use cases requiring reference and application ontologies and the use of real industry data, and using these experiences to encourage further development of BFO. We are also interested in how other ontologies perform and are active also in using the emerging ISO/CD TR 15926-14 ontology for maintenance data [16].

## 8. Conclusion

There is significant opportunity to use ontologies as an integral part of the data cleaning process for maintenance work order records. This paper focuses specifically on data quality checks on what maintenance activity was performed based on information contained in the MWO. Understanding these maintenance actions on a specific piece of equipment is analogous to a doctor being able to determine all the procedures and prescriptions a patient has had from historical medical records. In doing so, we cannot afford to rely only on the words used by the data generator to describe the *activity*. There are too many different people involved over the years in generating these records and they come from a wide range of backgrounds. As a result, to confirm the activity that was actually performed, engineers have to check if these *activity* words are consistent with data in the other fields of the MWO record. The ontology in this paper performs this check. The results are promising with the reasoning able to identify several data quality issues in the MWOs. The challenges for the ontology in making a final distinction between classes such as replace or repair and adjust or diagnose for corrective work order types are the same as for the engineers. In both cases additional information is required. The elucidations for each of the seven maintenance activity terms also reveal shortcomings, from an ontology perspective, with the descriptions for some of these *activity* terms in both the ISO 14224 and ISO 15926-4 standards.

The generation of instance data for the reasoning in the application-level ontology is dependent on having a suitable NLP pipeline. Work in NLP for MWOs is an active area of research. While the pipeline used here is state of the art for extracting *activity* and *item* information, we expect additional instance data and relational data to be available soon. This will assist with refining the reasoning of the application-level ontology presented in this paper. The combination of both NLP and ontologies for an industrial use case is an additional contribution to the literature for this paper. Further a number of the modular ontologies created for this project are re-usable for other use cases, namely the work order ontology, functional breakdown pump ontology, asset list ontology, and maintenance activity reference ontology.

The ontology community has a tendency to regard processes as failures if they perform in a less than perfect way. In the case of improvements in data quality of MWOs the task is so huge and laborious and the consequences of poor data quality so difficult to quantify and yet significant that any automation and standardisation is a worthwhile improvement. This work is a stepping stone to the use of ontologies for cleaning and processing maintenance work orders in industry. A necessary step in the digitisation journey modern industry players are now on.

## References

[1] M. Bertolini, D. Mezzogori, M. Neroni and F. Zammori, Machine learning for industrial applications: A comprehensive literature review, *Expert Systems with Applications* **175** (2021), 114820. doi:10.1016/j.eswa.2021.114820.

[2] T. Bikaun, T. French, M. Hodkiewicz, M. Stewart and W. Liu, LexiClean: An annotation tool for rapid multi-task lexical normalisation, in: *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, Association for Computational Linguistics, Online and Punta Cana, Dominican Republic, 2021, pp. 212–219, https://aclanthology.org/2021.emnlp-demo. doi:10.18653/v1/2021.emnlp-demo.25.

[3] T. Bikaun and M. Hodkiewicz, Semi-automated estimation of reliability measures from maintenance work order records, in: *PHMME 2021: Proceedings of the 6th European Conference of the Prognostics and Health Management Society 2021*, P. Do, S. King and O. Fink, eds, Vol. 6, The PHM Society, Turin, Italy, 2021, pp. 9–19.

[4] S. Borgo, M. Carrara, P. Garbacz and P.E. Vermaas, A formal ontological perspective on the behaviors and functions of technical artifacts, *Artificial Intelligence for Engineering Design, Analysis and Manufacturing (AI EDAM)* **23**(1) (2009), 3–21. doi:10.1017/S0890060409000079.

[5] S. Borgo, M. Carrara, P. Garbacz and P.E. Vermaas, Formalizations of functions within the DOLCE ontology, in: *Proceedings of the Eighth International Symposium on Tools and Methods of Competitive Engineering TMCE*, I. Horvath, F. Mandorli and Z. Rusak, eds, Vol. 1, Delft University of Technology, Netherlands, 2010, pp. 113–126.

[6] H. Brink, A. Krych, O.R. Cardenas and S. Tiwari, Establishing the right analytics-based maintenance strategy, McKinsey White Paper, 2021.

[7] M.P. Brundage, T. Sexton, M. Hodkiewicz, A. Dima and S. Lukens, Technical language processing: Unlocking maintenance knowledge, *Manufacturing Letters* **27** (2021), 42–46. doi:10.1016/j.mfglet.2020.11.001.

[8] M.P. Brundage, T. Sexton, M. Hodkiewicz, K.C. Morris, J. Arinez, F. Ameri, J. Ni and G. Xiao, Where do we start? Guidance for technology implementation in maintenance management for manufacturing, *Journal of Manufacturing Science and Engineering* **141**(9) (2019), 091005. doi:10.1115/1.4044105.

[9]  N. Chilamkurti, T. Torabi and R. Elhdad, Ontology-based framework for maintenance activity analysis and support: A case study for petroleum plant, *International Journal of System Assurance Engineering and Management* **5**(1) (2014), 84–98.

[10]  A. Conte, C. Bolland, L. Phan, M. Brundage and T. Sexton, The impact of data quality on maintenance work order analysis: A case study in historical HVAC maintenance work orders, in: *PHMME 2021: Proceedings of the 6th European Conference of the Prognostics and Health Management Society 2021*, P. Do, S. King and O. Fink, eds, Vol. 6, The PHM Society, Turin, Italy, 2021, pp. 11.

[11]  H. Deutsch and P. Garbacz, Relative identity, in: *The Stanford Encyclopedia of Philosophy, Winter 2022 Edn*, E.N. Zalta and U. Nodelman, eds, 2022, Metaphysics Research Lab, Stanford University.

[12]  A. Dima, S. Lukens, M. Hodkiewicz, T. Sexton and M.P. Brundage, Adapting natural language processing for technical text, *Applied AI Letters* **2**(3) (2021), 33. doi:10.1002/ail2.33.

[13]  Y. Gao, C. Woods, W. Liu, T. French and M. Hodkiewicz, Pipeline for machine reading of unstructured maintenance work order records, in: *Proceedings of the 30th. European Safety and Reliability Conference and 15th. Probabilistic Safety Assessment and Management Conference*, Venice, Italy, 2020.

[14]  N. Guarino, Artefactual systems, missing component and replaceability, in: *Artefact Kinds: Ontology and the Human-Made World*, M. Franssen, P. Kroes, T.A.C. Reydon and P.E. Vermaas, eds, Synthese Library, Vol. 365, Springer International Publishing, Cham, 2014, pp. 191–206.

[15]  M. Hodkiewicz and M.T.-W. Ho, Cleaning historical maintenance work order data for reliability analysis, *Journal of Quality in Maintenance Engineering* **22** (2016), 146–463. doi:10.1108/JQME-04-2015-0013.

[16]  M. Hodkiewicz, J.W. Klüwer, C. Woods, T. Smoker and T. French, An ontology for reasoning over engineering textual data stored in FMEA spreadsheet tables, *Computers in Industry* **131** (2021). doi:10.1016/j.compind.2021.103496.

[17]  M. Hodkiewicz, S. Lukens, M.P. Brundage and T. Sexton, Rethinking maintenance terminology for an industry 4.0 future, *International Journal of Prognostics and Health Management* **12**(1) (2021). doi:10.36001/ijphm.2021.v12i1.2932.

[18]  M. Hodkiewicz and N. Montgomery, Data fitness for purpose: Assessing the quality of industrial data for use in mathematical models, in: *8th International Conference on Modelling in Industrial Maintenance and Reliability*, Institute of Mathematics and Its Applications, Oxford, 2014, pp. 125–130.

[19]  M. Hodkiewicz, C. Woods, E. Low and F. Ameri, Towards a reference ontology for maintenance work management, in: *Proceedings of Interoperability for Enterprise Systems and Applications Workshops Co-Located with 10th International Conference on Interoperability for Enterprise Systems and Applications (I-ESA 2020)*, M. Zelm, B. Young, G. Doumeingts, H. Karray and L. Elmhadbhi, eds, CEUR Workshop Proceedings, Vol. 2900, CEUR-WS.org, Tarbes, France, 2020.

[20]  IEC, Dependability management – Part 3-14: Application guide – maintenance and maintenance support, Standard, IEC 60300-3-14:2004, International Electrotechnical Commission, Geneva, Switzerland, 2004.

[21]  IEC, Failure modes and effects analysis (FMEA and FMECA), Standard, IEC 60812:2018, International Electrotechnical Commission, Geneva, Switzerland, 2018.

[22]  Industrial Ontology Foundry, Industrial ontology foundry (IOF) core, 2021, Accessed: 2022-02-15. https://github.com/NCOR-US/IOF-BFO/tree/IOF-Core-2020.

[23]  ISO, ISO 14224: Petroleum, petrochemical and natural gas industries – collection and exchange of reliability and maintenance data for equipment, Standard, ISO14224:2016, International Organization for Standardization, Geneva, Switzerland, 2016.

[24]  ISO, ISO 15926-4:2019 Industrial automation systems and integration – integration of life-cycle data for process plants including oil and gas production facilities – Part 4: Initial reference data, Technical Report, ISO, Geneva, Switzerland, 2019.

[25]  ISO 15926-14: Industrial automation systems and integration – integration of life-cycle data for process plants including oil and gas production facilities – Part 14: Data model adopted for OWL 2 Direct Semantics, Community Draft, International Organization for Standardization, 2019.

[26]  ISO/IEC, ISO/IEC 21838-2:2021 Information technology – top-level ontologies (TLO) – Part 1: Requirements, Technical Report, ISO/IEC, Geneva, Switzerland, 2021.

[27]  ISO/IEC, ISO/IEC 21838-2:2021 Information technology – top-level ontologies (TLO) – Part 2: Basic Formal Ontology (BFO), Technical Report, ISO/IEC, Geneva, Switzerland, 2021.

[28]  A.T. James, O. Gandhi and S. Deshmukh, Knowledge management of automobile system failures through development of failure knowledge ontology from maintenance experience, *Journal of Advances in Management Research* (2017).

[29]  A. Jordan, M. Selway, W. Mayer, G. Grossmann and M. Stumptner, An ontological core for conformance checking in the engineering life-cycle, in: *Formal Ontology in Information Systems – Proceedings of the Eighth International Conference, FOIS 2014*, P. Garbacz and O. Kutz, eds, Frontiers in Artificial Intelligence and Applications, Vol. 267, IOS Press, Rio de Janeiro, 2014, pp. 358–371.

[30]  M.H. Karray, F. Ameri, M. Hodkiewicz and T. Louge, ROMAIN: Towards a BFO compliant reference ontology for industrial maintenance, *Applied Ontology* **14**(2) (2019), 155–177. doi:10.3233/AO-190208.

[31]  M.H. Karray, B. Chebel-Morello and N. Zerhouni, A formal ontology for industrial maintenance, *Applied Ontology* **7**(3) (2012), 269–310. doi:10.3233/AO-2012-0112.

[32]  D.P. Lupp, M. Hodkiewicz and M.G. Skjæveland, Template libraries for industrial asset maintenance: A methodology for scalable and maintainable ontologies, in: *Proceedings of the 12th International Workshop on Scalable Semantic Web Knowledge Base Systems Co-Located with 19th International Semantic Web Conference (ISWC 2020)*, T. Liebig, A. Fokoue and Z. Wu, eds, CEUR Workshop Proceedings, Vol. 2757, CEUR-WS.org, 2020, pp. 49–64, Technical University of Aachen.

[33]  M. Madhikermi, S. Kubler, J. Robert, A. Buda and K. Främling, Data quality assessment of maintenance reporting procedures, *Expert Systems with Applications* **63** (2016), 145–164. doi:10.1016/j.eswa.2016.06.043.

[34] C. Masolo, L. Vieu, E. Bottazzi, C. Catenacci, R. Ferrario, A. Gangemi and N. Guarino, Social roles and their descriptions, in: *Principles of Knowledge Representation and Reasoning: Proceedings of the Ninth International Conference (KR2004)*, D. Dubois, C.A. Welty and M. Williams, eds, AAAI Press, 2004, pp. 267–277.

[35] A. Matsokis, H.M. Karray, B. Chebel-Morello and D. Kiritsis, An ontology-based model for providing semantic maintenance, *IFAC Proceedings Volumes* **43**(3) (2010), 12–17. doi:10.3182/20100701-2-PT-4012.00004.

[36] R. Mizoguchi, K. Kozaki and Y. Kitamura, Ontological analyses of roles, in: *2012 Federated Conference on Computer Science and Information Systems (FedCSIS)*, M. Ganzha, L.A. Maciaszek and M. Paprzycki, eds, IEEE Computer Society, 2012, pp. 489–496.

[37] R. Molina, K. Unsworth, M. Hodkiewicz and E. Adriasola, Are managerial pressure, technological control and intrinsic motivation effective in improving data quality?, *Reliability Engineering & System Safety* **119** (2013), 26–34. doi:10.1016/j.ress.2013.04.009.

[38] J.J. Montero Jiménez, R. Vingerhoeds, B. Grabot and S. Schwartz, An ontology model for maintenance strategy selection and assessment, *Journal of Intelligent Manufacturing* (2021), 1–19.

[39] J. Moubray, *Reliability-Centered Maintenance*, Butterworth-Heinemann, 1997.

[40] G.D. Murphy, Improving the quality of manually acquired data: Applying the theory of planned behaviour to data quality, *Reliability Engineering & System Safety* **94**(12) (2009), 1881–1886. doi:10.1016/j.ress.2009.05.008.

[41] L. Nagy, T. Ruppert and J. Abonyi, Ontology-based analysis of manufacturing processes: Lessons learned from the case study of wire harness production, *Complexity* **2021** (2021).

[42] M. Navinchandran, M.E. Sharp, M.P. Brundage and T.B. Sexton, Discovering critical KPI factors from natural language in maintenance work orders, *Journal of Intelligent Manufacturing* (2021), 1–19.

[43] M.V. Ottermo, S. Håbrekke, S. Hauge and L. Bodsberg, Technical language processing for efficient classification of failure events for safety critical equipment, in: *PHMME 2021: Proceedings of the 6th European Conference of the Prognostics and Health Management Society 2021*, P. Do, S. King and O. Fink, eds, Vol. 6, The PHM Society, Turin, Italy, 2021, pp. 9.

[44] A. Polenghi, I. Roda, M. Macchi and A. Pozzetti, Multi-attribute ontology-based criticality analysis of manufacturing assets for maintenance strategies planning, *IFAC-PapersOnLine* **54**(1) (2021), 55–60. doi:10.1016/j.ifacol.2021.08.192.

[45] M. Poveda-Villalón, A. Gómez-Pérez and M.C. Suárez-Figueroa, Oops!(ontology pitfall scanner!): An on-line tool for ontology evaluation, *International Journal on Semantic Web and Information Systems (IJSWIS)* **10**(2) (2014), 7–34. doi:10.4018/ijswis.2014040102.

[46] D. Rajpathak and R. Chougule, A generic ontology development framework for data integration and decision support in a distributed environment, *International Journal of Computer Integrated Manufacturing* **24**(2) (2011), 154–170. doi:10.1080/0951192X.2010.531291.

[47] D. Rajpathak, H. Siva Subramania and P. Bandyopadhyay, Ontology-driven data collection and validation framework for the diagnosis of vehicle health management, *International Journal of Computer Integrated Manufacturing* **25**(9) (2012), 774–789. doi:10.1080/0951192X.2012.665187.

[48] D. Rajpathak, Y. Xu and I. Gibbs, An integrated framework for automatic ontology learning from unstructured repair text data for effective fault detection and isolation in automotive domain, *Computers in Industry* **123** (2020), 103338. doi:10.1016/j.compind.2020.103338.

[49] D.G. Rajpathak, An ontology based text mining system for knowledge discovery from the diagnosis data in the automotive domain, *Computers in Industry* **64**(5) (2013), 565–580. doi:10.1016/j.compind.2013.03.001.

[50] M. Rausand and K. Øien, The basic concepts of failure analysis, *Reliability Engineering & System Safety* **53**(1) (1996), 73–83. doi:10.1016/0951-8320(96)00010-5.

[51] SAE, Evaluation criteria for reliability-centered maintenance (RCM) processes, Standard, JA1011-2009, SAE International, 2009.

[52] SAE International, A guide to the reliability-centered maintenance (RCM) standard, Standard, SAE JA1012, SAE International, 2018.

[53] T. Sexton, M. Hodkiewicz, M.P. Brundage et al., Categorization errors for data entry in maintenance work-orders, in: *PHM 2019: Proceedings of the Annual Conference of the Prognostics and Health Management Society 2019*, N.S. Clements, ed., Vol. 11, The PHM Society, 2019.

[54] T. Sexton, M. Hodkiewicz, M.P. Brundage and T. Smoker, Benchmarking for keyword extraction methodologies in maintenance work orders, in: *PHM 2018: Proceedings of the Annual Conference of the Prognostics and Health Management Society 2018*, A. Bregon and M. Orchard, eds, Vol. 10, The PHM Society, 2018.

[55] T.B. Sexton, M.P. Brundage et al., Nestor: A tool for natural language annotation of short texts, *Journal of Research of the National Institute of Standards and Technology* **124** (2019). doi:10.6028/jres.124.029.

[56] J. Sikorska, M. Hodkiewicz and L. Ma, Prognostic modelling options for remaining useful life estimation by industry, *Mechanical Systems and Signal Processing* **25**(5) (2011), 1803–1836. doi:10.1016/j.ymssp.2010.11.018.

[57] J.I. Single, J. Schmidt and J. Denecke, Knowledge acquisition from chemical accident databases using an ontology-based method and natural language processing, *Safety Science* **129** (2020), 104747. doi:10.1016/j.ssci.2020.104747.

[58] SINTEF, Offshore reliability data handbook, Technical Report, Høvik: Det Norske Veritas, 2015.

[59] B. Smith, F. Ameri, H. Cheong, D. Kiritsis, D. Sormaz, C. Will and J.N. Otte, A first-order logic formalization of the industrial ontology foundry signature using basic formal ontology, in: *Proceedings of the Joint Ontology Workshops 2019 Episode V: The Styrian Autumn of Ontology – 10th International Workshop on Formal Ontologies Meet Industry (FOMI)*, A. Barton, S. Seppälä and D. Porello, eds, CEUR Workshop Proceedings, Vol. 2518, CEUR-WS.org, Graz, Austria, 2019.

[60] S. Stern, S. Farioli, E. Eisenschmidt, S. Morachioli, C. Kienzler, D. Jeske and N. Schaal, The future of maintenance for distributed assets, McKinsey White Paper, 2020.

[61] M. Stewart, M. Hodkiewicz, W. Liu and T. French, MWO2KG and Echidna: Constructing and exploring knowledge graphs from maintenance data, *Institution of Mechanical Engineers – Journal of Risk and Reliability* (2022), 1–13.

[62] M. Stewart, W. Liu and R. Cardell-Oliver, Redcoat: A collaborative annotation tool for hierarchical entity typing, in: *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP): System Demonstrations*, Association for Computational Linguistics, Hong Kong, China, 2019, pp. 193–198, https://aclanthology.org/D19-3033.

[63] M. Stewart, W. Liu, R. Cardell-Oliver and R. Wang, Short-text lexical normalisation on industrial log data, in: *2018 IEEE International Conference on Big Knowledge (ICBK)*, X. Wu, Y. Ong, C.C. Aggarwal and H. Chen, eds, IEEE Computer Society, 2018, pp. 113–122. doi:10.1109/ICBK.2018.00023.

[64] K. Unsworth, E. Adriasola, A. Johnston-Billings, A. Dmitrieva and M. Hodkiewicz, Goal hierarchy: Improving asset data quality by improving motivation, *Reliability Engineering & System Safety* **96**(11) (2011), 1474–1481. doi:10.1016/j.ress.2011.06.003.

[65] M. Uschold and M. Gruninger, Ontologies: Principles, methods and applications, *The Knowledge Engineering Review* **11**(2) (1996), 93–136. doi:10.1017/S0269888900007797.

[66] UWA NLP-TLP Group, The University of Western Australia: Natural & Technical Language Processing Group, 2021, Accessed: 2021-11-11. https://nlp-tlp.org/.

[67] B.A. Weiss, M.P. Brundage, Y. Tamm, T. Makila, J. Pellegrino et al., Summary report on the industry forum for monitoring, diagnostics, and prognostics for manufacturing operations, NIST Advanced Manufacturing Series 100-23, 2019.

[68] C. Woods, M. Selway, M. Hodkiewicz, F. Ameri, M. Stumptner and W. Sobel, On the notion of maintenance state for industrial assets, in: *Proceedings of the Joint Ontology Workshops 2021 Episode VII: The Bolzano Summer of Knowledge Co-Located with the 12th International Conference on Formal Ontology in Information Systems (FOIS 2021), and the 12th International Conference on Biomedical Ontologies (ICBO 2021)*, E.M. Sanfilippo, O. Kutz, N. Troquard, T. Hahmann, C. Masolo et al., eds, CEUR Workshop Proceedings, Vol. 2969, CEUR-WS.org, 2021, Rheinisch-Westfaelische Technische Hochschule Aachen, Lehrstuhl Informatik V.