



# Using formal thesauri and controlled vocabulary as the interface between the unstructured data and axiomatic ontologies

**Farhad Ameri**

NIST TLP COI Event

April 2021, Gaithersburg, MD

# High-level Research questions

---

**How Knowledge Organization Systems (KOS) can be used in an integrated fashion to support formal knowledge extraction from unstructured data?**

- Supply Chain Use Case :
  - Data type: Manufacturing Capability data
  - Data source: company website
- Maintenance Use Case :
  - Data type: MWO records
  - Data Source: CMMS

# Maintenance Use Case Objectives

---

- To expose and formalize the semantics of unstructured maintenance data.
- To visualize MWO records as RDF knowledge graphs
- To use RDF knowledge graphs for maintenance diagnostics and root-cause analysis.

# Knowledge Organization Systems (KOS)

Controlled  
vocabulary for  
Information  
Retrieval

## KOS Types

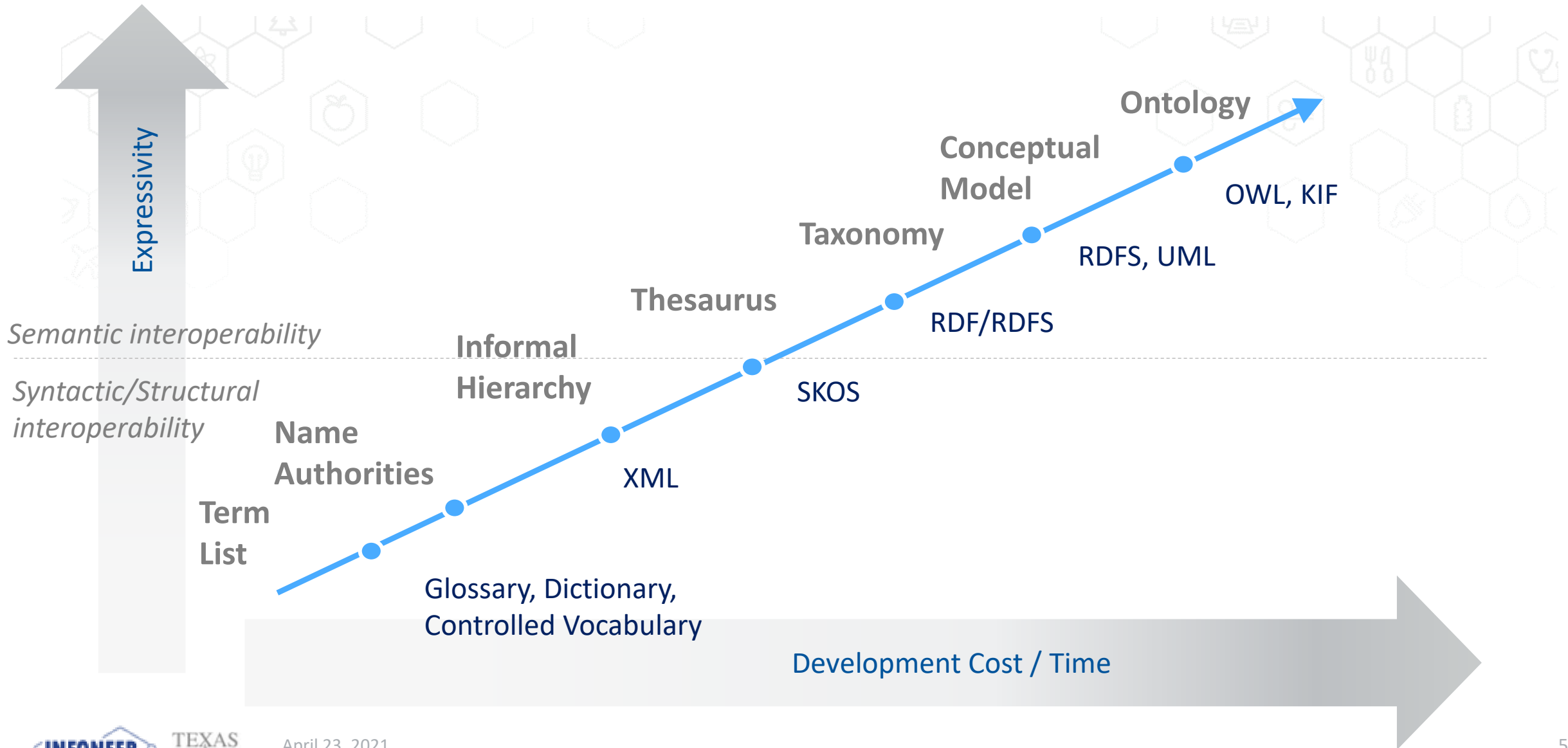
term lists  
synonym rings  
name authorities  
taxonomies  
**thesauri**  
glossaries  
dictionaries  
gazetteers  
categorization schemes  
classification systems  
subject heading schemes  
semantic networks  
**ontologies**

- Knowledge organization system is any system of terms or **scheme** that is created to organize, manage, and retrieve information.
- They vary in complexity, structure, function, and expressivity.
- Services:
  - Abstraction & Indexing
  - Tagging
  - Information retrieval and term disambiguation
  - Navigation and translation
  - Reasoning and inference

Source: Introduction to taxonomies and other knowledge organization systems, Heather Hedden and Helmut Nagy, Semantic Web Company

April 23, 2021

# Semantic Spectrum of Knowledge Organization Systems



# Thesaurus

---

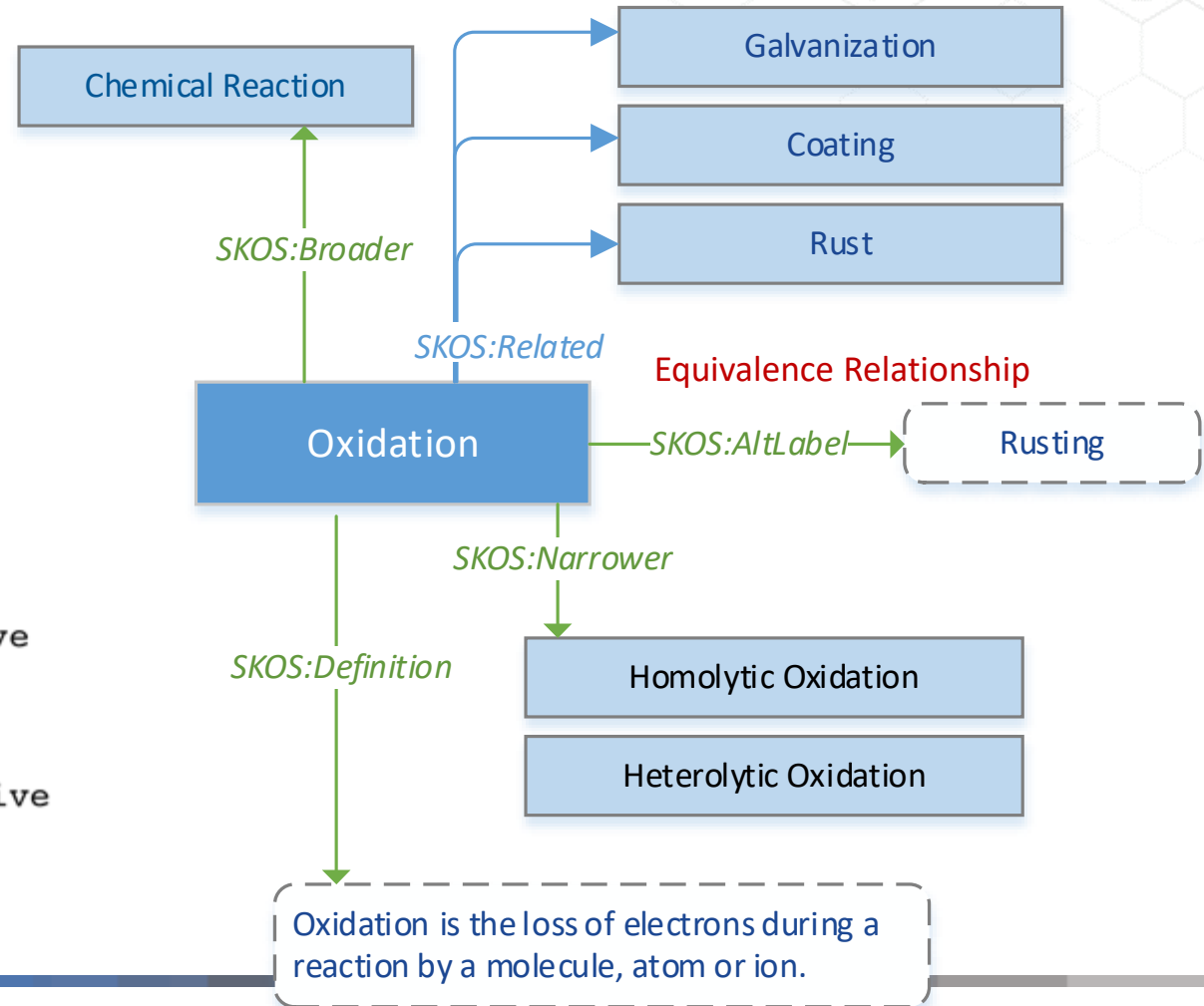
- A controlled vocabulary that represents three type of relationships between terms or concepts:
  - **Hierarchical:** broader term/narrower terms
  - **Associative:** related terms
  - **Equivalence:** preferred label /alternative label
- Created in accordance with standards:
  - ISO 25964 (2011, 2013) and ANSI/NISO Z39.19

# SKOS model (Simple Knowledge Organization System)

SKOS is a **standard** data model for representation of knowledge organization systems.

- W3C Standard
- Web-native syntax
- Encoded using XML/RDF
- Each concept has a unique URI
- Machine-readable
- Simple semantics

```
skos:semanticRelation
├── skos:related
├── skos:broaderTransitive
│   └── skos:broader
├── skos:narrowerTransitive
│   └── skos:narrower
```





# Why Thesaurus?

## Thesaurus

They can accommodate large or constantly growing vocabulary

They can be used to directly tag unstructured data

They are easier to develop and extend compared to ontologies

We want to use the best of both worlds

## Ontology

Logic-based semantics

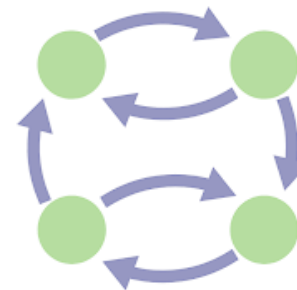
High level expressivity and formality



## Thesaurus

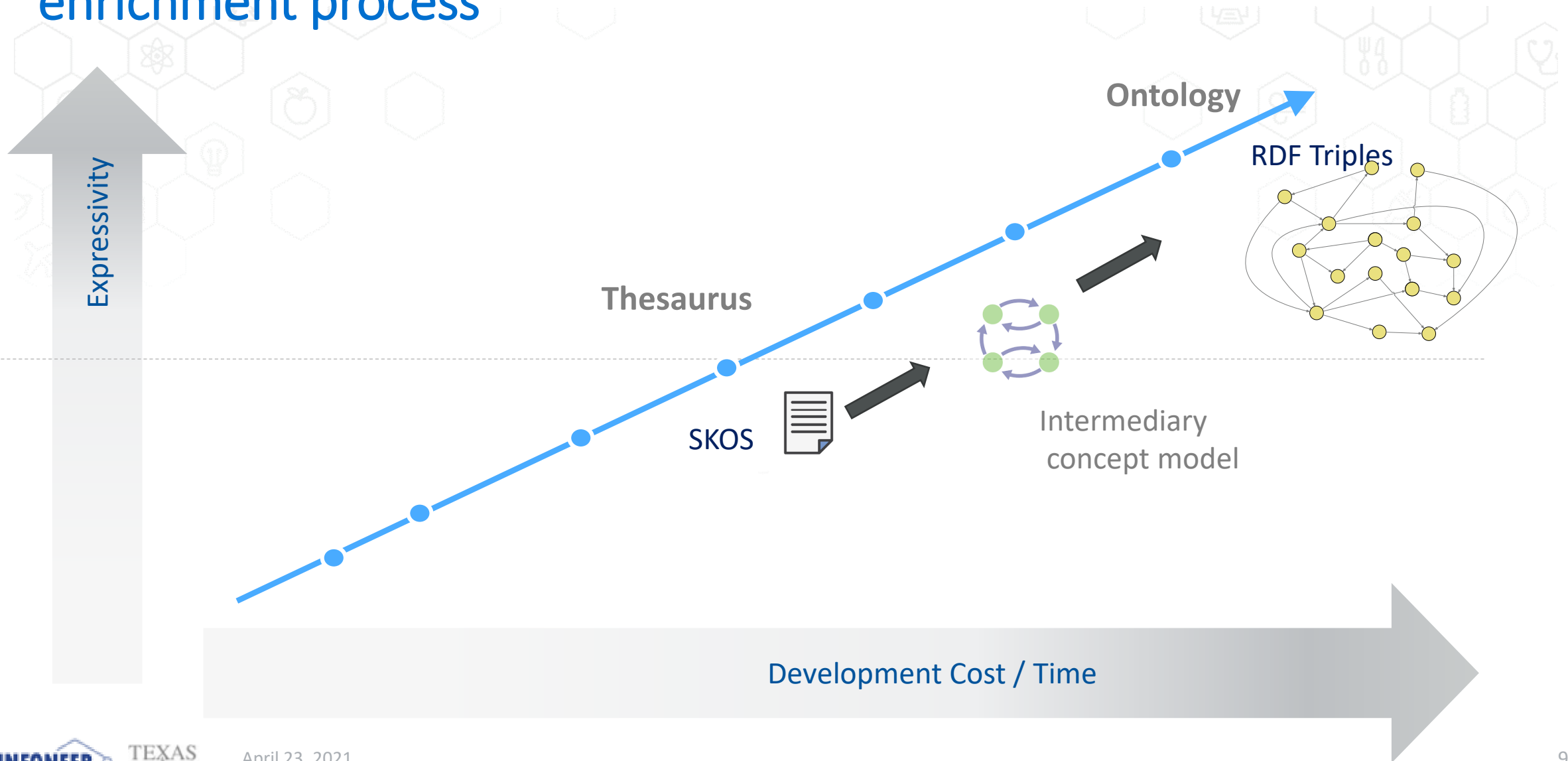
Lexical semantics

Ease of extension





# Overall Approach: Incremental structuring and semantic enrichment process



# Example

Raw text (portion of reality) from a MWO Record

Solenoid valve on upper lift cyl leaking. Hyd pressure reduced on Boom. Technician called. valve replaced, works normal

Explicit Entities

Implicit Entities

Artifact

States

Failure

Events

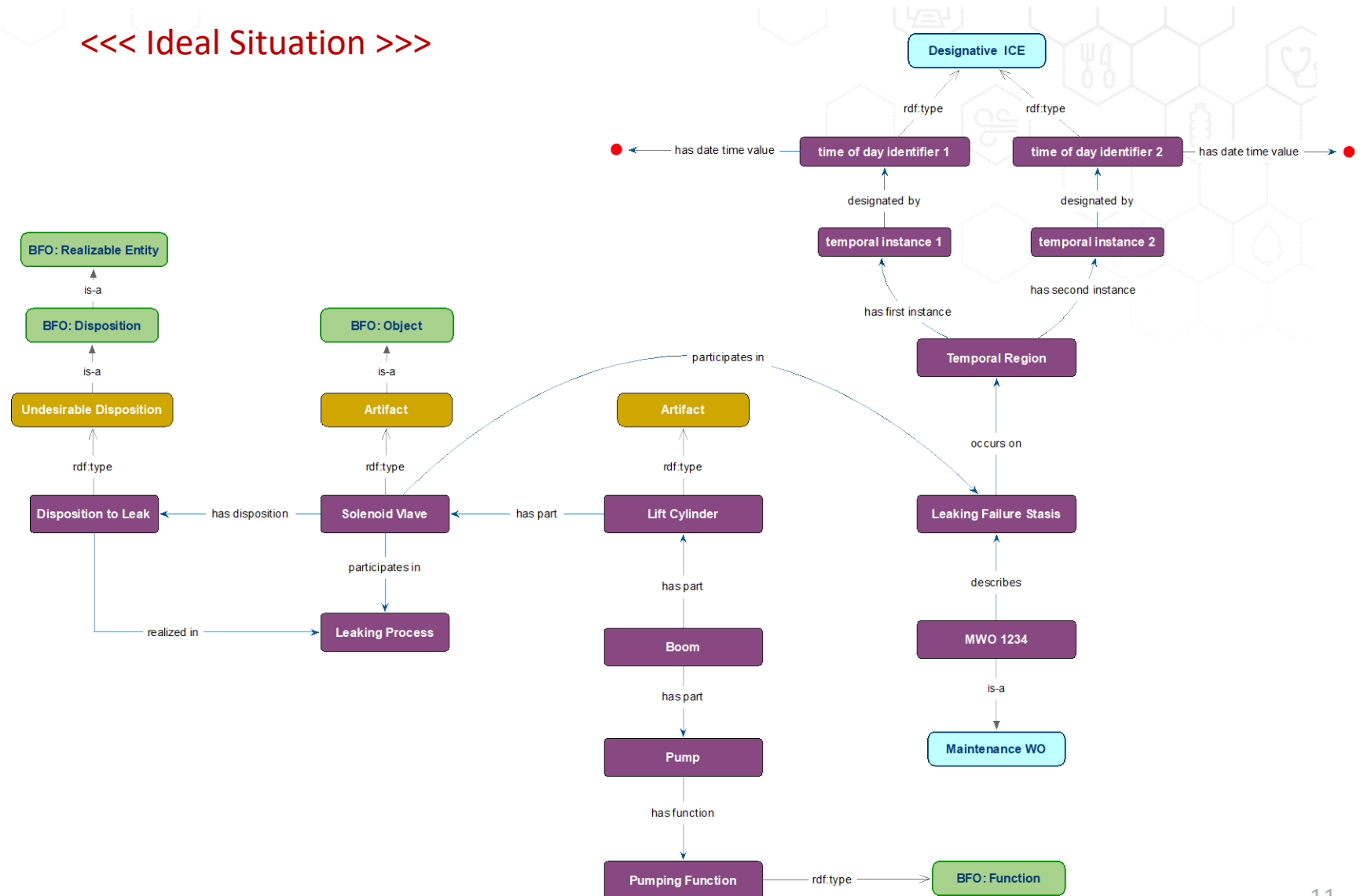
Agent

Process

# Ontological Representation of problem-solution description

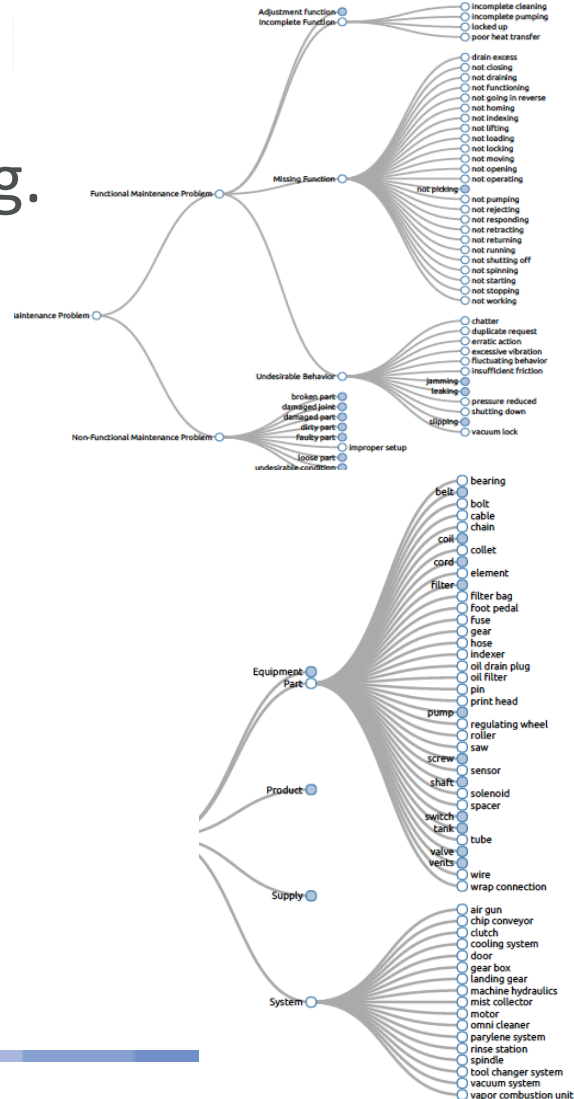
Solenoid valve on upper lift cyl leaking. Hyd pressure reduced on Boom. Technician called. valve replaced, works normal

<<< Ideal Situation >>>

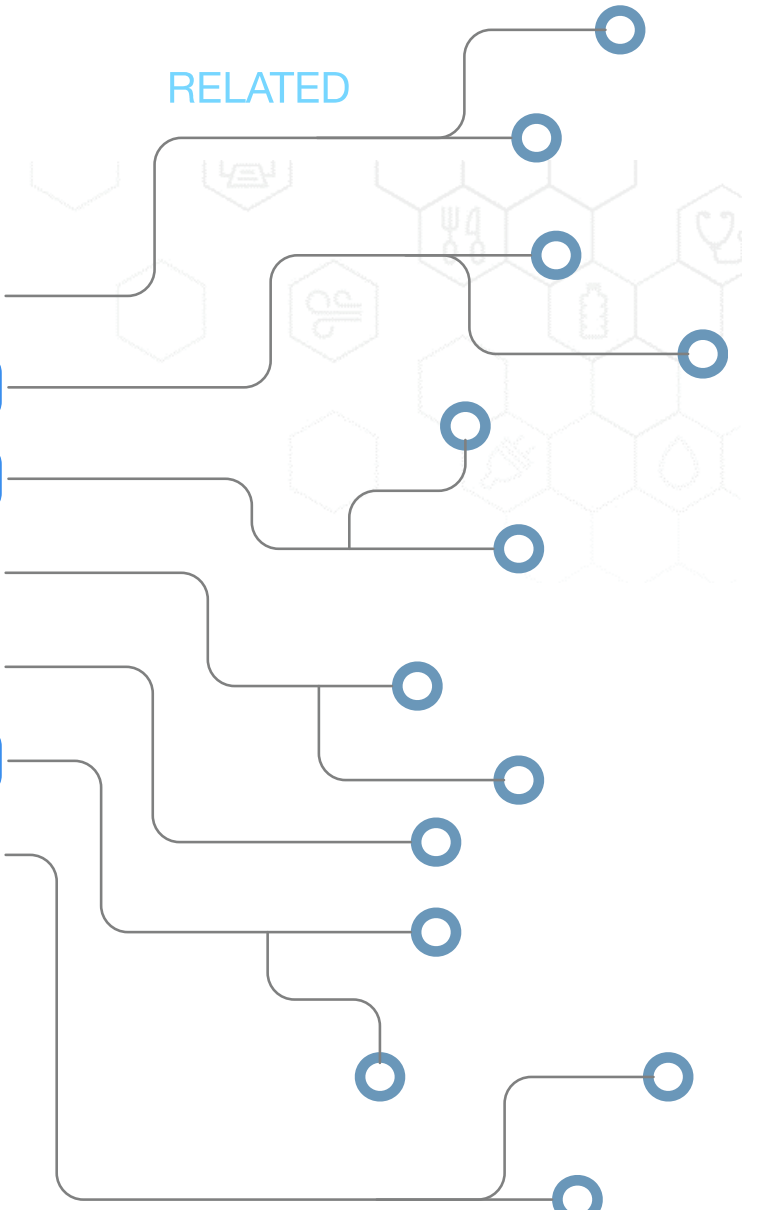


# 2-step process: First annotate the text with SKOS concepts

Solenoid valve on upper lift cyl leaking. Hyd pressure reduced on Boom. Technician called. valve replaced works normal

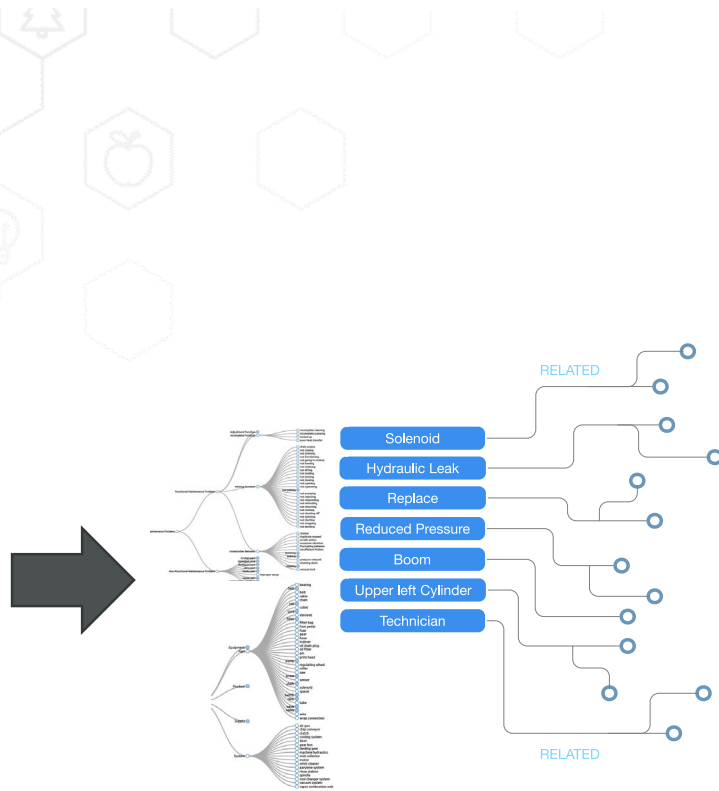


- Solenoid
- Hydraulic Leak
- Replace
- Reduced Pressure
- Boom
- Upper left Cylinder
- Technician

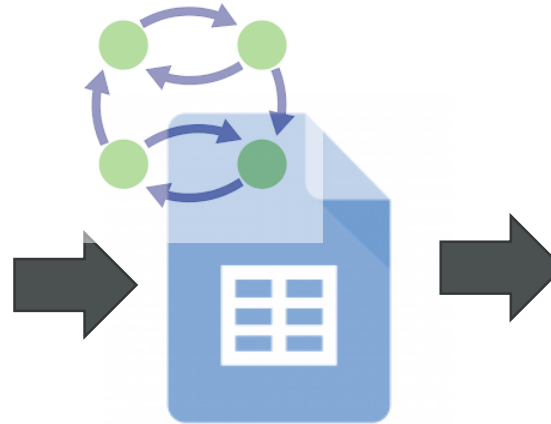


# Create an intermediary representation (structured)

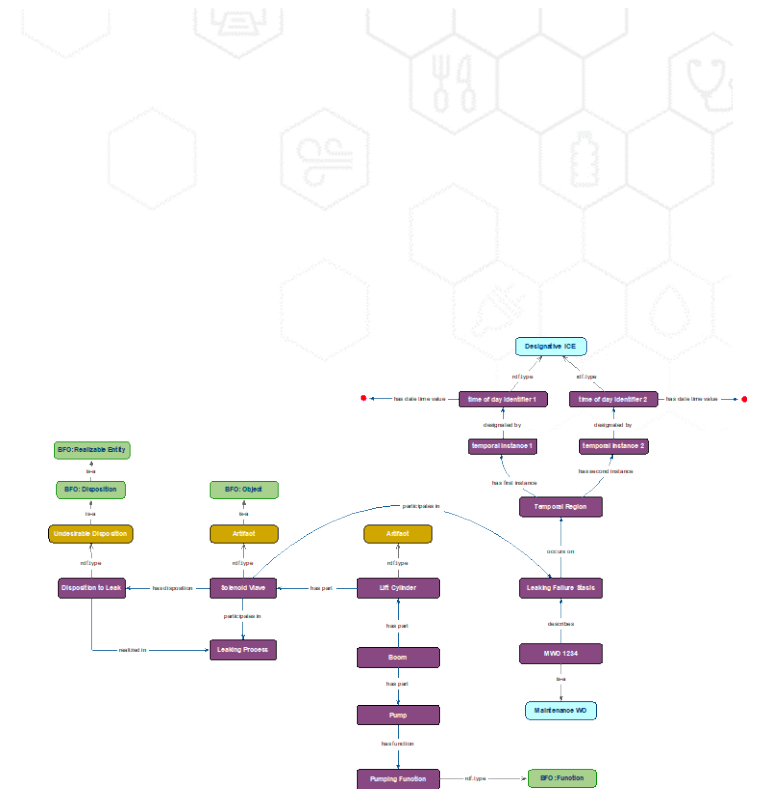
Solenoid valve on upper lift cyl leaking. Hyd pressure reduced on Boom. Technician called. valve replaced works normal



## Intermediary representation (Concept Model)



- *Graph*
- *Table*

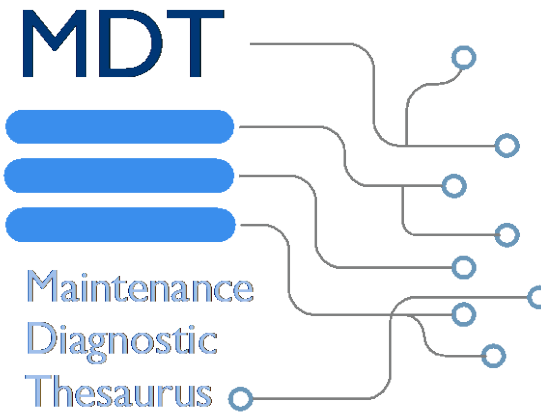


## Semantic Web Zone

SKOS, RFD/XML, RDFS. OWL

# Maintenance Diagnostics Thesaurus

For this purpose, we developed Maintenance Diagnostics Thesaurus or MDT

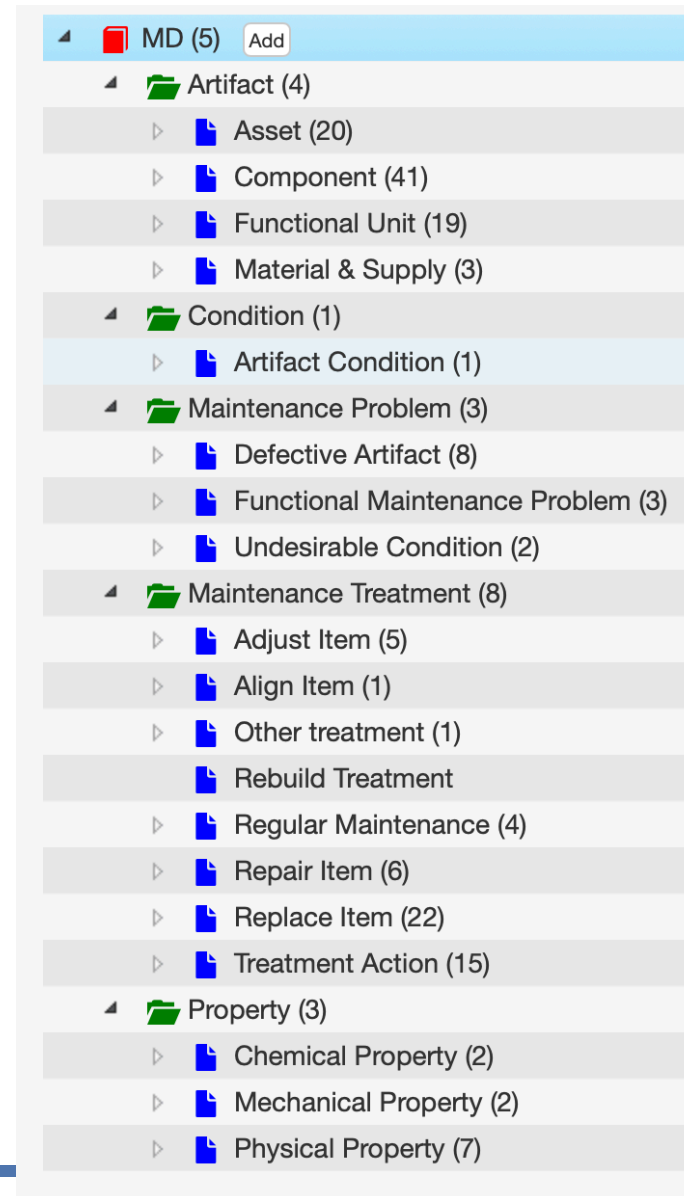


A SKOS model for representation of controlled vocabulary in the maintenance domain.

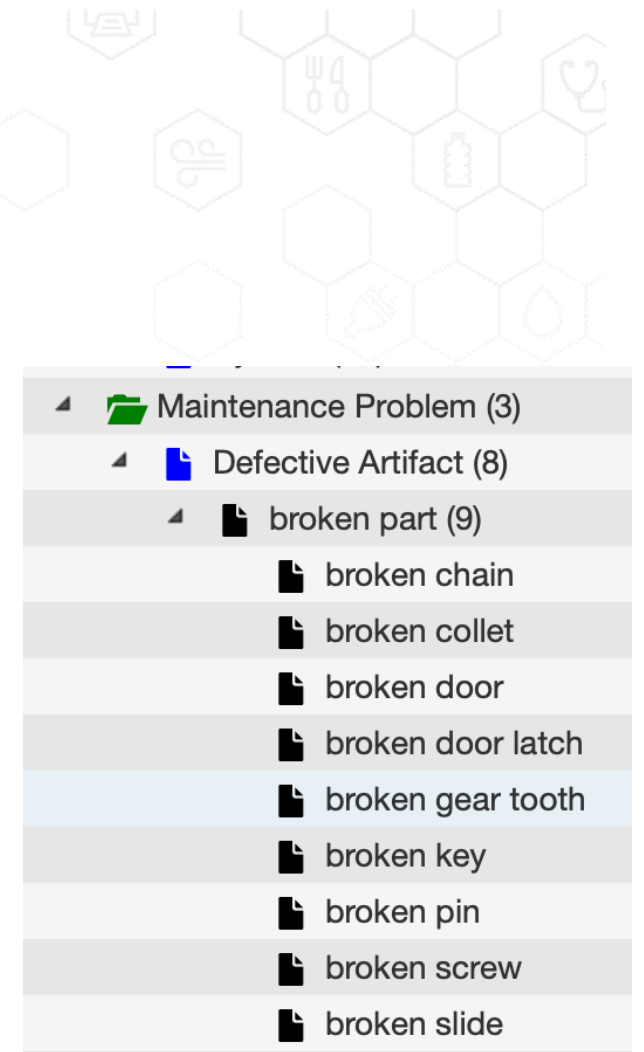
# The scope of Maintenance Diagnostic Thesaurus (MDT)

## Scope of MDT:

- Failures
- Treatment
- Quality (color, temperature, viscosity, lumen)
- Feature (crack, hole, breakage)
- States
  - Working, not working
- Process
  - Malfunctioning (e.g., leaking, rotating slowly)
    - Including not functioning or reduced functioning
- Items
  - Component
  - Equipment/Machine
  - Sub-system



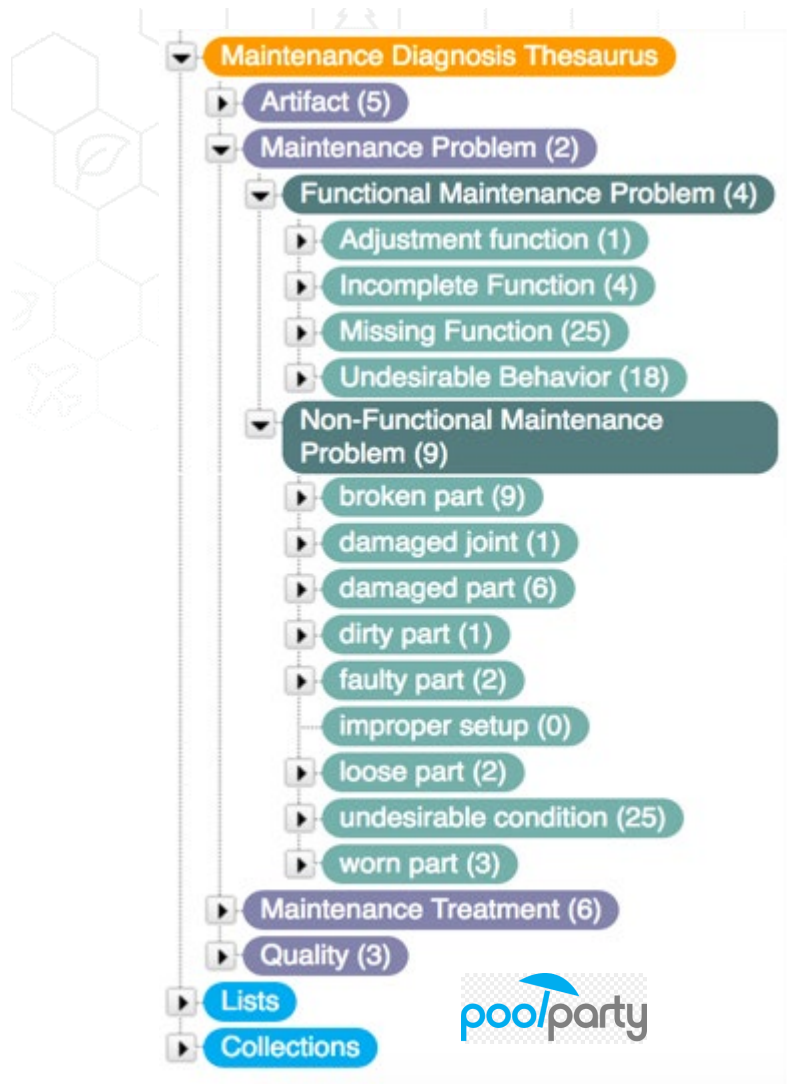
```
graph TD; MD["MD (5) Add"] --> Artifact["Artifact (4)"]; MD --> Condition["Condition (1)"]; MD --> MaintenanceProblem["Maintenance Problem (3)"]; MD --> MaintenanceTreatment["Maintenance Treatment (8)"]; MD --> Property["Property (3)"]; Artifact --> Asset["Asset (20)"]; Artifact --> Component["Component (41)"]; Artifact --> FunctionalUnit["Functional Unit (19)"]; Artifact --> MaterialSupply["Material & Supply (3)"]; Condition --> ArtifactCondition["Artifact Condition (1)"]; MaintenanceProblem --> DefectiveArtifact["Defective Artifact (8)"]; MaintenanceProblem --> FunctionalMaintenanceProblem["Functional Maintenance Problem (3)"]; MaintenanceProblem --> UndesirableCondition["Undesirable Condition (2)"]; MaintenanceTreatment --> AdjustItem["Adjust Item (5)"]; MaintenanceTreatment --> AlignItem["Align Item (1)"]; MaintenanceTreatment --> OtherTreatment["Other treatment (1)"]; MaintenanceTreatment --> RebuildTreatment["Rebuild Treatment"]; MaintenanceTreatment --> RegularMaintenance["Regular Maintenance (4)"]; MaintenanceTreatment --> RepairItem["Repair Item (6)"]; MaintenanceTreatment --> ReplaceItem["Replace Item (22)"]; MaintenanceTreatment --> TreatmentAction["Treatment Action (15)"]; Property --> ChemicalProperty["Chemical Property (2)"]; Property --> MechanicalProperty["Mechanical Property (2)"]; Property --> PhysicalProperty["Physical Property (7)"];
```



```
graph TD; DefectiveArtifact["Defective Artifact (8)"] --> BrokenPart["broken part (9)"]; BrokenPart --> BrokenChain["broken chain"]; BrokenPart --> BrokenCollet["broken collet"]; BrokenPart --> BrokenDoor["broken door"]; BrokenPart --> BrokenDoorLatch["broken door latch"]; BrokenPart --> BrokenGearTooth["broken gear tooth"]; BrokenPart --> BrokenKey["broken key"]; BrokenPart --> BrokenPin["broken pin"]; BrokenPart --> BrokenScrew["broken screw"]; BrokenPart --> BrokenSlide["broken slide"];
```



# MD Thesaurus



**SKOS TOOL**

■ = thesaurus  
■ = concept scheme  
■ = top concept  
■ = normal concept

Label search:

- ▲ ■ Maintenance Diagnosis Thesaurus (5)
  - ▲ ■ Artifact (5)
    - ▶ ■ Equipment (18)
    - ▶ ■ Part (37)
    - ▶ ■ Product (1)
    - ▶ ■ Supply (3)
    - ▶ ■ System (18)
  - ▲ ■ Maintenance Problem (2)
    - ▶ ■ Functional Maintenance Problem (3)
    - ▶ ■ Non-Functional Maintenance Problem (10)
  - ▲ ■ Maintenance Treatment (8)
    - ▲ ■ Adjustment Treatment (5)
      - adjust airspeed
      - adjust hinges
      - adjust print head
      - adjust roller
      - adjust sensor
    - ▶ ■ Alignment Treatment (1)
    - ▶ ■ Other treatment (1)
    - ▶ ■ Rebuild Treatment
    - ▶ ■ Regular Maintenance (4)

Top-level Thesaurus

## Thesaurus statistics

Total nodes: 350

Total concept schemes: 5

Total concepts: 345

Total preferred labels: 345

Total alternative labels: 101

Total hidden labels: 3

# Entity Extractor

- Home
- Corpus Analyzer \*TODO
- Thesaurus Manager
- Entity Extractor
- CM Builder
- CM Manager
- Capability Scorer
- Contact

## Upload text

Paste text  URL (case sensitive)  Upload CSV file

B49D4E42-8994-46D8-8641-000234CDA205 GEN NOT WORKING, KEN W/ GENIE ADVISED TO CHANGE RESISTOR AND REPLACE POT SWITCH INSIDE BOX. REPLACED PARTS AS ADVISED AND RESET NEW REOSTATE. UNIT NOW WORKING. GWS

Upload CSV file. First row must be a header row (column descriptions).

no file selected

Select thesaurus:

Maintenance Diagnosis Thesaurus

- Include zero-occurrence concepts
- Include top-level concepts
- Analyze CSV file rows separately
- Show URL preview page

URL depth:  0

SKOS TOOL

# Detected Concepts

Detected as skos:prefLabel

Detected as skos:altLabel

## Analysis results

B49D4E42-8994-46D8-8641-000234CDA205 GEN NOT WORKING, KEN W/ GENIE ADVISE  
D TO CHANGE RESISTOR AND REPLACE POT SWITCH INSIDE BOX.  
REPLACED PARTS AS ADVISED AND RESET NEW REOSTATE. UNIT NOW WORKING.  
GWS

Export Text

Word count: 29

Include URLs in exported text

SKOS TOOL

# Exported Concept Model for the MWO record



Sort table:

Alphabetical  Occurrences

Concept (preferred label)	Occurrences
change	2
generator	1
not working	1
replace	2
reset	1
resistor	1
switch	1
unit working	1

Export type:

2-column  4-column

Export Table



Exported Table

Concept Schema	Top Concept	Concept (preferred label)	Occurrences
Artifact	System	generator	1
Maintenance Problem	Functional Maintenance	not working	1
Maintenance Treatment	Treatment Action	change	2
Artifact	Part	resistor	1
Maintenance Treatment	Treatment Action	replace	2
Artifact	Part	switch	1
Maintenance Treatment	Treatment Action	reset	1
Problem Resolved	Resolution Term	unit working	1

Formats available:

- CSV
- RDF/JSON

# Thesaurus Development challenges

- To adequately annotated all WO records even in a single company, we need a large thesaurus.
- Thesaurus extension is a bottleneck
- Mainly manual process
  - One concept at a time
  - Batch import

Select terms:

COMPARTMENT DOOR LATCH BROKEN NEEDS LATCH  
OIL LEAK FROM JOYSTICK. DEFECTIVE SEALS ON JOYSTICK.  
Refurb: 9/12/13 GENIE wont start when cold-35 degrees , unit came with no glow plugs  
Breakage & Break it. TAPE  
will not charge. found bad charger  
HYD LEAK HYD MOTOR FOR GENERATOR LEAKING  
muffler clamp broke off allowing muffler to move around and dig in weldment  
bottom of boom gouging, wear blocks and shims falling out. block spacer gouging boom.  
WILL NOT START/DO ANYTHING. FOUND BATTERY IS FAULTY  
Wearpads were cracked.  
door latch came apart  
Unit will not operate/Faulty contactor  
Brake on crank failed.Faulty mechanism.

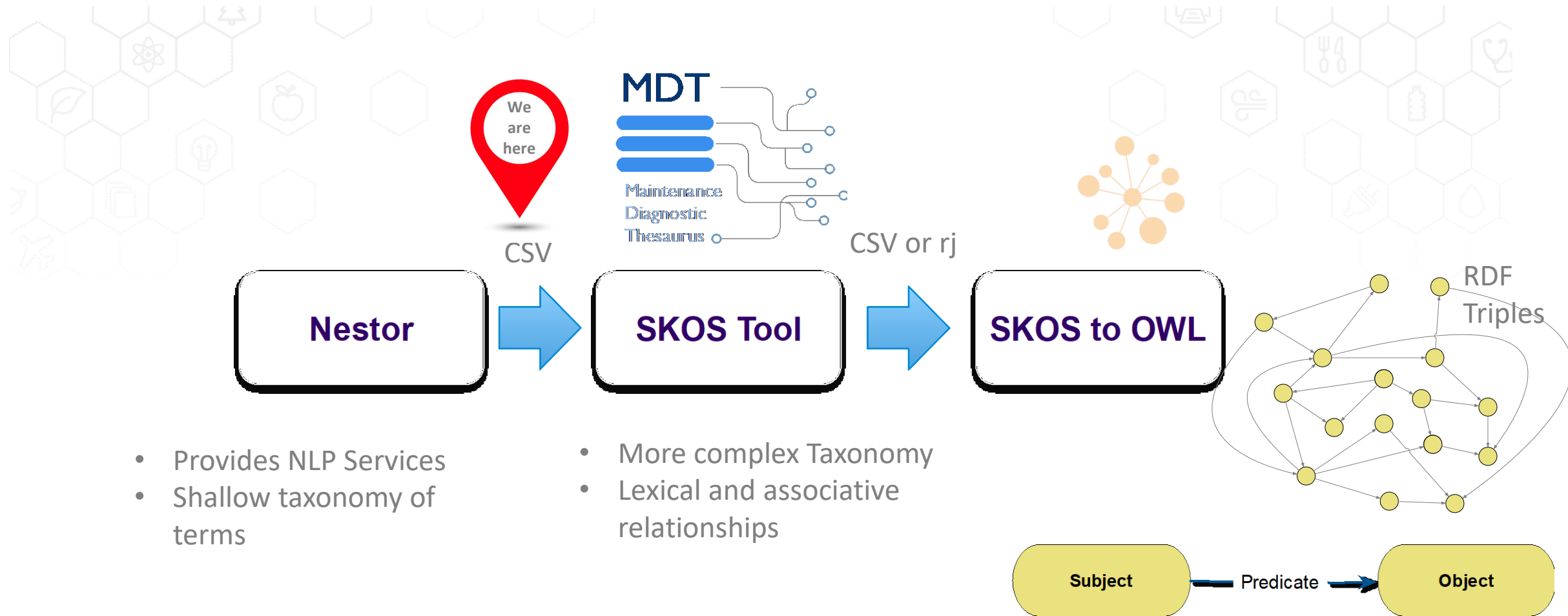
New concept:

Parent concept:

aluminum housing  
housing  
JOINT  
steel bolt  
**Functional Unit**  
contactor  
mechanism  
WINCH  
unit

**SKOS TOOL**

# Semi-automated Extension of MDT



# Categorizing Items (from Nestor) in MDT

## Nestor Output (Item column)

	I	
0	battery, unit	
1	charger, unit	-
2	boss, boss_fitting, motor, oil, ring	
3	box, genie, pot, resistor, switch, unit	
4	engine, fuse	
5	limit_switch, switch, unit	
6	exhaust, flange, gasket, tube	l
7	board	F
8	oil, ring, unit	C
9	control_valve, oil, ring, valve	-

Candidate  
Terms



## Categorize new concepts

Please select the parent of each new concept. Concepts with the same group number will be related to each other (enter 0 for no relations).

Group:	<input type="text" value="0"/>	Pref. label:	<input type="text" value="control valve"/>	Select parent:	<input type="text" value="--"/>	<input type="checkbox"/>	2
Group:	<input type="text" value="0"/>	Pref. label:	<input type="text" value="ring"/>	Select parent:	<input type="text" value="Component"/>	<input type="checkbox"/>	2
Group:	<input type="text" value="0"/>	Pref. label:	<input type="text" value="board"/>	Select parent:	<input type="text" value="Component"/>	<input type="checkbox"/>	2
Group:	<input type="text" value="0"/>	Pref. label:	<input type="text" value="exhaust"/>	Select parent:	<input type="text" value="Component"/>	<input type="checkbox"/>	2
Group:	<input type="text" value="0"/>	Pref. label:	<input type="text" value="limit switch"/>	Select parent:	<input type="text" value="Component"/>	<input type="checkbox"/>	2
Group:	<input type="text" value="0"/>	Pref. label:	<input type="text" value="engine,"/>	Select parent:	<input type="text" value="Functional U"/>	<input type="checkbox"/>	2
Group:	<input type="text" value="0"/>	Pref. label:	<input type="text" value="box"/>	Select parent:	<input type="text" value="Component"/>	<input type="checkbox"/>	2

No relations

Submit >



# Next Steps

---

- Extending the thesaurus and ontology in tandem
  - Better integration with Nestor
  - More diverse dataset is needed to feed the thesaurus extension process
- Tool Development :
  - SKOS to OWL : Receives SKOS Concept Model and converts it to RDF graph
    - Reasoning to extract inferred knowledge (hidden / implicit)
  - Diagnosis Tool: Provides diagnostic support by querying the RDF graph