PO.ID 102

Making Sense of Wind-farm Data

A novel semantic framework for interrogating heterogeneous wind-farm data

Frank O'Connor M.Sc., Dr. Paul Leahy, Dr. Richard Kavanagh, Dr. Des Farren
ServusNet Informatics/University College Cork (UCC)

ServusNet knowledge is power

Introduction

Systems which manage wind farms typically generate large volumes of data and managing this data is particularly problematic for three main reasons.

Firstly, issues are often reported in ambiguous or unintuitive language. They are rarely classified, and when they are, it is usually to a coarse proprietary classification system which is of little use when counting issues, particularly when tabulating issues across multiple vendors.[1]

Secondly, a loose coupling may exist between the data produced by machines (such as alarms and events) and data produced by humans (such as maintenance and service reports) which leads to heterogeneous data. It is not uncommon for records of work undertaken to maintain or improve assets to be poorly recorded, e.g. as unstructured written reports. This means that attempting to automatically analyse data for correlations between alarms and events, and maintenance activities is difficult if not impossible.

Thirdly, there is a wealth of information around the boundaries of wind farms that have a significant role in helping to place machine data in the correct context, not just information about the assets' locations like longitude/latitude, etc., but also geographical facts like nearby obstructions or topographical features. This local information often only exists in the heads of experienced experts, and can be difficult to formally capture.

Objectives

Proprietary → Open/Standard

Heterogeneous → Homogeneous

Context poor → Context rich

Methods

The first stage of the methodology used Unified Modelling Language (UML*1) to capture and describe a total system information model for a wind farm. The model is composed of a set of UML diagrams.

1. **Use case diagrams** identify all of the systems, people and activities that produce or consume wind farm data. (see figure 1)
2. **A Package diagram** shows high-level interactions between systems, people and data, with particular attention to information boundaries. (see figure 2)
3. **Class diagrams** show detailed wind farm information model, with particular attention to data attributes and information composition. (see figure 3)

An initial draft of the model was created using material from published literature[2][3][4]. Subsequent versions were reviewed with industry practitioners including operators and experts from international standards groups.

The second stage in the methodology created a semantic model based on the UML model (see figure 4.) An open semantic language called OWL*2 was selected to codify the wind farm model. The two main steps in this process were:

1. Identify appropriate existing ontologies
2. Develop new ontologies and taxonomies where necessary

*1 http://www.uml.org/

*2 http://www.w3.org/TR/owl-semantics/

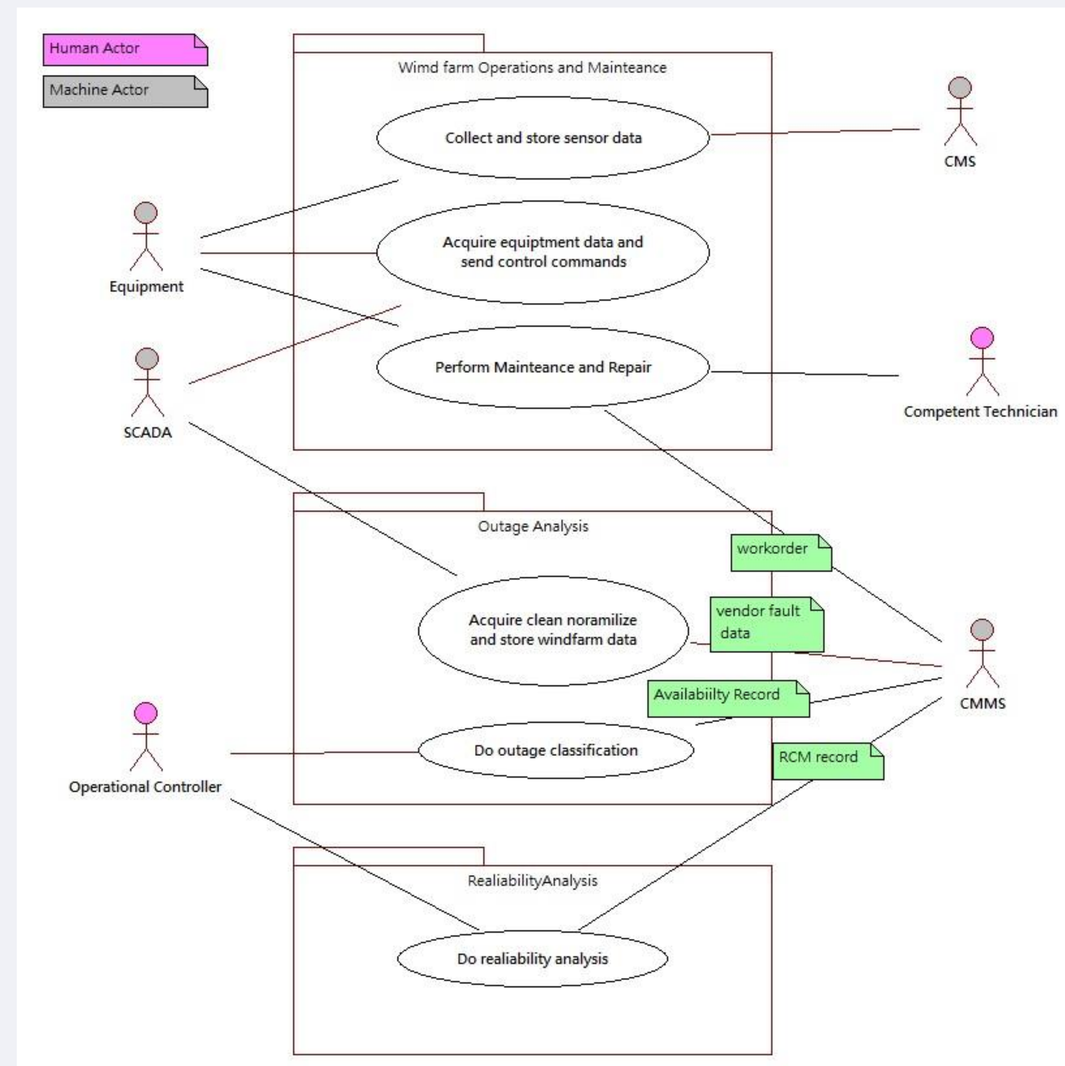

Figure 1: High level wind farm activities

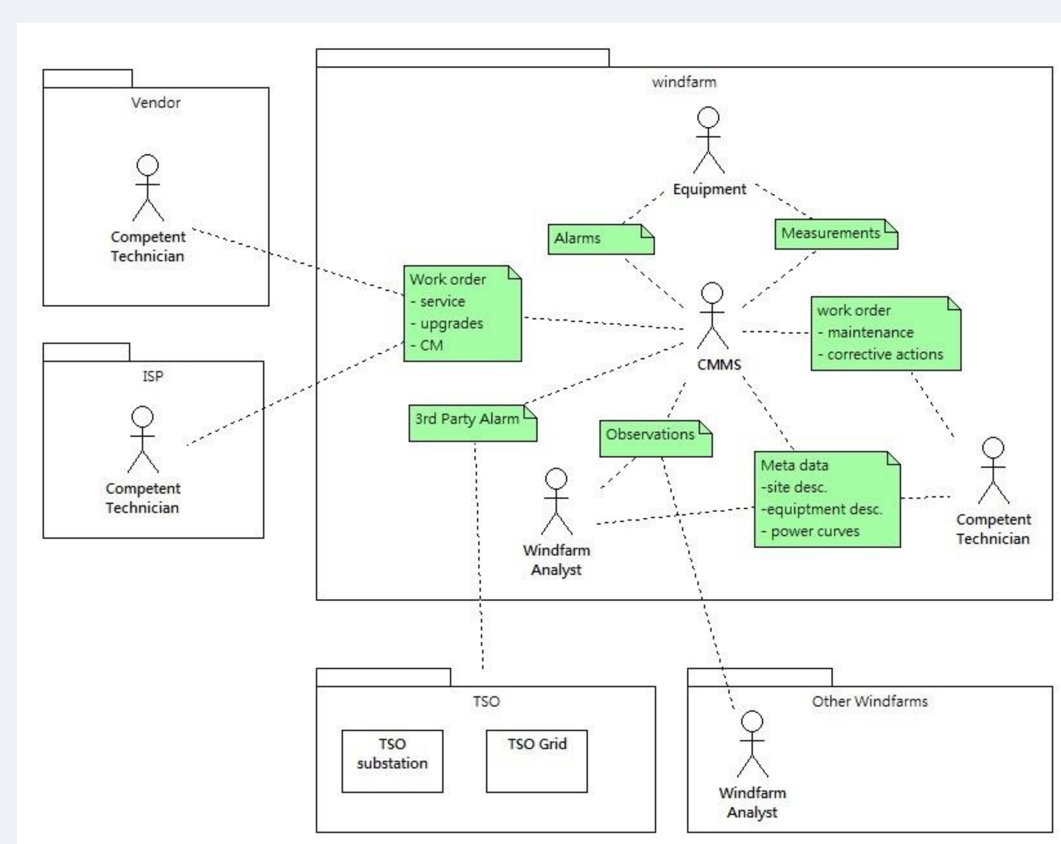

Figure 2: –High level wind farm data flows

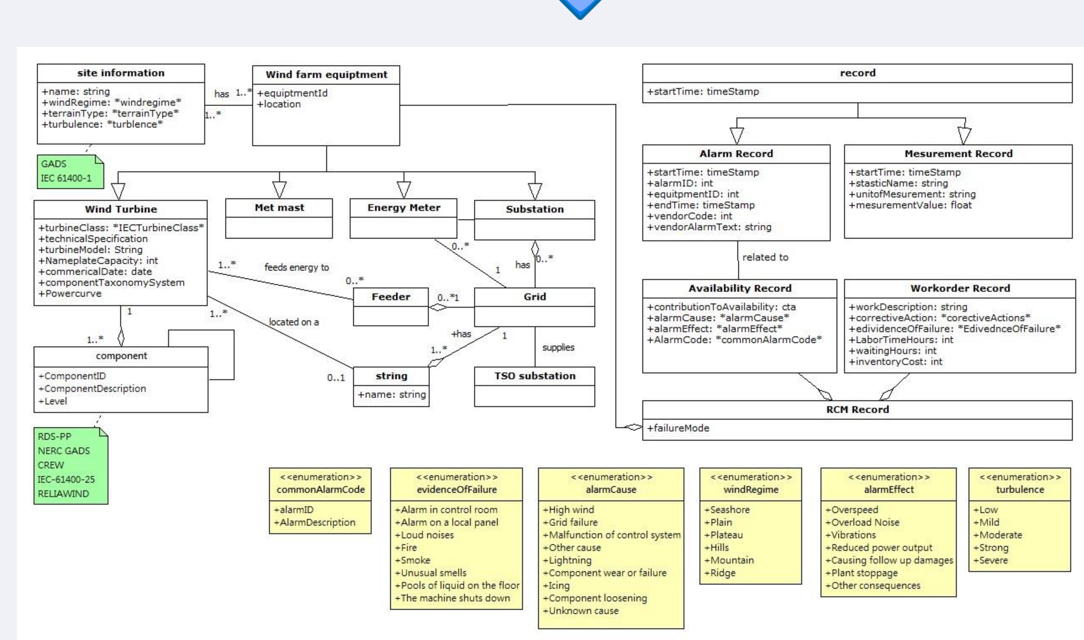

Figure 3 – Detailed wind farm information model

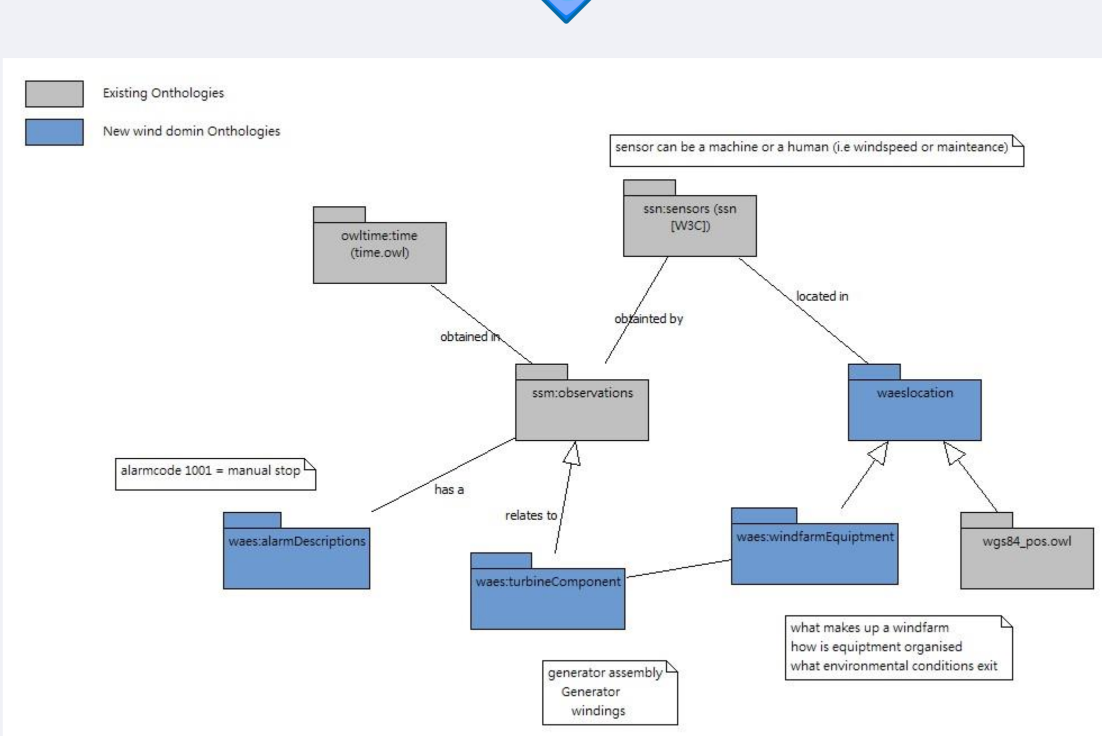

Figure 4 –Wind farm semantic model

For more information about this project please visit
http://www.frankwoconnor.com/WAES/
For comments and feedback on the wind farm model please visit
http://www.frankwoconnor.com/WAES/EWEA2015/feedback

Results

An agent was created to generate statements about wind farm data in the form of subject–predicate–object expressions, known as triples. The subject denotes the resource, and the predicate denotes a relationship between the subject and the object. The object can be a property such as a value or another resource. An example of a triple graph, created from the wind farm semantic model, is shown in figure 5 below.

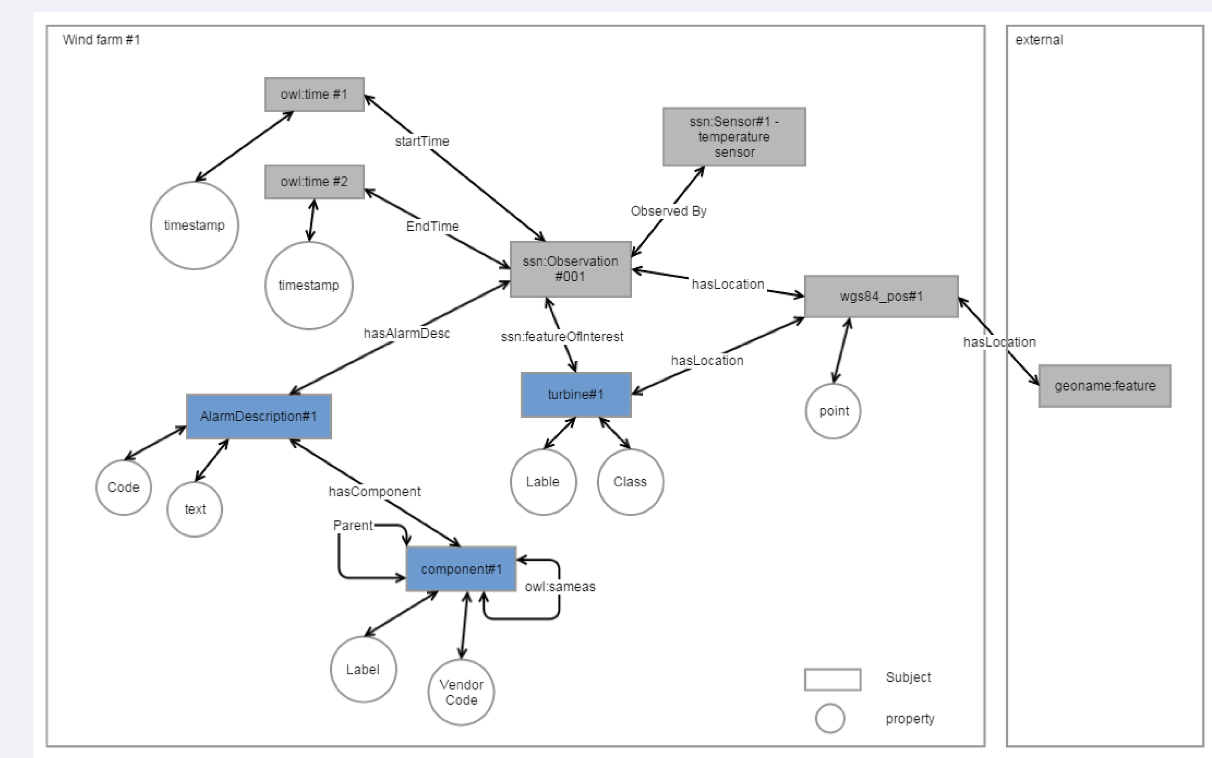

Figure 5: Semantic graph fragment

The set of triples generated by the agent is persisted in a triple store, a purpose-built database for the storage and retrieval of triples through semantic queries.

1. Heterogeneous wind farm data is converted to a simple yet flexible homogeneous schema. Any wind farm fact can be expressed as an observation, for example a SCADA pressure sensor measurement or a met mast temperature sensor reading. Even facts like noting an equipment failure, a software upgrade or a part replacement can be captured and described using the observation/sensor pattern
2. The implementation also has facts about wind farm alarms. In the graph fragment presented in figure 5, it can be seen that alarm resources have code and text objects as well a relationship to a component resource. This component resource represents a non–proprietary enumeration of component types and is typically drawn from appropriate industry standards e.g. IEC 61400-25. Hierarchical facts about components are easily accommodated using the 'hasParent' predicate.
3. Contextual data can be added to resources using the predicate object pattern. Semantic data supports the linking and sharing of data so, for example, when a resource such as a turbine is modelled, and this resource already exists in another data set (e.g. geonames) it is possible to simply reference this resource rather than recreate it. This can provide a wealth of contextual knowledge at no extra cost. E.g. find all turbines with alarm observations for blade components, where the turbines are within 10k of a road.

Conclusions

This poster has described some of the difficulties with managing wind farm data and identified the need for better information management solutions to help make sense of growing quantities of wind-farm data.

It has been demonstrated how a detailed windfarm information model was developed using a formal methodology which included a review of exiting windfarm information models as well as inputs from domain experts. A high level overview of the proposed wind farm semantic model was shown as well as a diagram detailing how a fact, such as, an alarm can be represented as a generic observation using semantic graph notation.

A software implementation based on the described semantic model was also shown. This implementation addressed the three issues identified with wind farm data. An approach was described where proprietary data classifications can be normalized to an open wind farm taxonomy. It was shown how heterogeneous data can be represented in a simple yet flexible homogeneous pattern, and finally it was shown how data can be linked to other independent sets of domain knowledge, and how this can provide important and significant contextual information at little or no extra cost.

References

[1] M. Melik-Merkumians, a Zoitl, and T. Moser, "Ontology-based fault diagnosis for industrial control applications," *2010 IEEE 15th Conf. Emerg. Technol. Fact. Autom. (ETFA 2010)*, pp. 1–4, Sep. 2010.

[2] R. Isermann, "Supervision, fault-detection and fault-diagnosis methods — An introduction," *Control Eng. Pract.*, vol. 5, no. 5, pp. 639–652, May 1997.

[3] S. Faulstich, "Optimising maintenance data to boot turbine Efficiency," *DEWEK 2010*, 2010.

[4] Anders Johnsson, "How to make information in the maintenance process more useful? ", Nordic Windpower confrence Operations & Mainteance, May 2010, p5

Acknowledgments

I would like to acknowledge the support of the Irish Research Council (IRC), for funding this research, my two supervisors, Dr. Paul Leahy and Dr. Richard Kavanagh for their advice and guidance, and my employment partner ServusNet informatics for their continued support.