ELEMENT 9 **INFORMATION MANAGEMENT**

Harmonization of Pump Schemas with the ISO 15926 Reference Data Library

December 2011

Fiatech™

Harmonization of Pump Schemas with the ISO 15926 Reference Data Library

December 2011

Acknowledgements

Fiatech has produced this report at the request of and for the benefit of its members and in cooperation with the Hydraulic Institute. We wish to acknowledge the contributions of those individuals whose efforts and input have significantly influenced this publication.

Notably, this report was produced by a project team that voluntarily carried out the objective to provide a technical document for the capital projects industry. This team was led by Mark Palmer, NIST, and consisted of: Darius Kanga, Bechtel; Robin Benjamins, Bechtel; Manoj Dharwadkar, Bentley; and Russell Adams, RHAdams Consulting.

I wish to acknowledge the special efforts of Trygve Dahl and Scott Frees of Intelliquip, who served as consultants to the project and were the primary authors of this publication. Additionally, Thomas (Rusty) Haggard provided copyediting services, and Dallas Peters, Dallas Peters Design, assisted in graphic design production.

Among the staff, Fiatech Project Manager Sharon Bickford contributed significantly to the development of this publication and her efforts are appreciated.

Publication of this report is made possible through the contributions by Fiatech members.

Raymond E. Topping, PE

Director, Fiatech

Preface

Fiatech (*www.fiatech.org*) is an industry-led consortium, housed at The University of Texas at Austin, that provides global leadership in identifying and accelerating the development, demonstration, and deployment of emerging technologies and innovative practices to deliver the highest business value throughout the life cycle of all types of capital projects.

Fiatech is member-driven and is comprised of approximately 85 companies and partner organizations that include owners and operators from the industrial, power, and retail markets; leading providers of engineering, design, and construction services; software and equipment suppliers and technology providers; research institutes; and universities. Fiatech is a clearinghouse for innovative ideas where members can quickly learn about new processes, methods, and materials, but it also collectively funds and executes development, demonstration, and deployment projects. Project teams are formed to identify and accelerate the adoption of technologies and systems; demonstration projects are conducted to validate and perfect new approaches or methods; and teams are formed to aid and facilitate the deployment of those breakthrough initiatives that have been identified.

This publication was developed at the request of Fiatech members and in cooperation with the Hydraulic Institute for the benefit of the entire capital projects industry.

Table of Contents

Introduction

This document completes the deliverable for Tasks 1, 2, and 3 defined under the AEX/ISO 15926 Harmonization initiative as part of the Engineering Equipment Life-Cycle Application Tools (EELCAT) project. An interim progress report dated April 29,2011, is also available.

1.1 Objective

The objective of this initiative is to provide exploratory analysis of the work required to harmonize three technical activities: ISO 15926, AEX cfiXML, and the Hydraulic Institute's EDE 50.7 standard. Each of these activities endeavor to facilitate efficient exchange of capital facility and engineering data in electronic format. However, they approach and solve these issues in different manners.

1.1.1 AEX cfiXML

The Fiatech AEX cfiXML schemas were developed to automate information exchange for the design, procurement, delivery, installation, and operation and maintenance of engineered equipment. The schemas define the precise format and information organization of an XML document to be used to transfer equipment data between parties. The schemas encompass an extremely broad domain space, supporting many equipment types and use cases. To support information exchange between parties, each party must have a software implementation capable of reading and writing XML documents that adheres to the AEX schema. The technical objective of working with the XML instance documents is supported by off-the-shelf software tools. The core exercise when exchanging information using AEX is a process of locating the field(s) one wishes to transmit/store within the AEX schema and then creating the application code to perform the task. When exchanging information between two parties, the parties must be in agreement concerning the proper "location" (XPath) for the desired data.

1.1.2 HI-EDE 50.7

The AEX Schema is the result of combining the requirements for many types of equipment and many different organizations and viewpoints. Individual equipment engineering applications only need a subset of this dataset for most data transaction use cases, however, and a negotiated agreement on the bounded set of required fields (and how to represent them within the AEX Schema) can at times be difficult to realize, particularly between disparate third parties. Recognizing this, the Hydraulic Institute (HI) set out to publish a standard "data dictionary" enumerating the common data elements transacted in a variety of usage scenarios involving pumping equipment.

The initial focus of the work was centrifugal, rotary, and vertically suspended pumps, resulting in the *Hydraulic Institute Standard HI 50.7-2010 for Electronic Data Exchange for Pumping Equipment*, or HI-EDE 50.7. The data dictionary enumerates data fields and provides definitions and cross-listings into standard, well-known paper-based datasheets such as API 610 and ANSI/ASME B73. The HI-EDE does not define the mechanism or format of the electronic data exchange, only the semantic meanings of the data to be exchanged at individual points in the life cycle. However, it does include in its published listing for each data field an XPath supported by the AEX Schema. This mapping is intended to allow third parties to exchange AEX instance documents without embarking on their own individual mapping negotiation. The HI-EDE is being extended to cover additional pump types and can also provide mappings to alternative technical implementations in addition to AEX, such as the ISO 15926 RDL. However, only mappings to the AEX cfiXML schema currently exist.

1.1.3 ISO 15926

ISO 15926 aims to provide a model and library of classes and templates suitable for representing life-cycle information about technical installations and their components using semantic web principles[1]. The model

[1] ISO 15926 is technology neutral, in that a variety of technologies may be used to represent and transact data. Part 8 and Part 9 of the standard include requirements for the use of semantic web technologies such as RDF and SPARQL however.

library (Reference Data Library or RDL) is extensible such that parties can modify and augment its contents to suit the needs of many different industry domains. The RDL consists of "tiers" of localized libraries. Some of these tiers are maintained by individual organizations while others are more "standardized," the premise being that additions at the more localized tiers can be made without significant "vetting." However, organizations focusing on interoperability should restrict their usage to the more standardized, top-level tiers as much as possible. Exchanging data complying with the ISO 15926 RDL is accomplished using a variety of technologies ranging from semantic web languages (RDF, OWL), XML, and even spreadsheets. While the tools and framework for building an RDL suitable for full life-cycle information is mature, the RDL itself is currently incomplete and under significant development by members of the ISO 15926 community.

The ISO 15926 standard is broken into 10 parts[2], consisting of:

- Part 1 - Overview and Fundamental Principles
- Part 2 - Data Model
- Part 3 - Reference Data for Geometry and Topology
- Part 4 - Reference Data Classes
- Part 6 - Reference Data Additions
- Part 7 - Templates
- Part 8 - RDF/OWL Implementation Specification
- Part 9 - SPARQL Specification
- Part 10 - Abstract Test Methods
- Part 11 - Simplified Industrial Usage including Gellish

[2] Part 5: "Registration Procedure" has been removed from the standard and absorbed within other, more general parts of the standard.

For the purposes of this initiative, Parts 2, 4, and 7 are the most relevant as they govern what types of data and concepts are represented in the standard and how to define relationships between them. Parts 6 is somewhat relevant, as implementing the alignment; new classes may need to be created in the RDL. Parts 8-11 focus more on the implementation of ISO 15926 compliant software systems, which is not the primary focus of this alignment initiative.

1.1.3.1 iRING

iRING is a full implementation of the ISO 15926 standard. It was developed through a successfully demonstrated implementation project called "Camelot." iRING will be a key tool in the current harmonization initiative as it allows for easy access and browsing of the ISO 15926 RDL model/library.

1.1.3.2 EELCAT Project

Fiatech is sponsoring the Engineering Equipment Life-Cycle Application Tools (EELCAT) project. The focus of EELCAT is to develop application tools and data delivery specifications for the life cycle of engineered equipment from specification through procurement to system design and operation and maintenance while utilizing industry standards, i.e., ISO 15926 and AEX (as included in HI 50.7 and ISO 13709) for the exchange of data.

A "life cycle of engineered equipment" can be defined as the period of time from inception to disposal or when the equipment is no longer useful. EELCAT encompasses several initiatives across the Fiatech Capital Projects Technology Roadmap. It addresses the entire life cycle of engineered equipment rather than only the procurement phase, as was the case with the Fiatech project, "Automating Equipment Information Exchange (AEX). "

1.2 Task Descriptions

The three task activities defined in this initiative are summarized below:

- **Task 1**: Review current ISO 15026 Works in Progress (WIPs) RDL development tools and the associated process.
  - *Deliverable 1: Provide written report of findings.*
- **Task 2**: Meet with Fiatech members currently working in ISO 15926 WIPs and developing templates to review their results, assess the status of WIPs, and collect recommendations.
  - *Deliverable 2*: *Provide written meeting notes documenting discussions.*
- **Task 3**: Assess and document the alignment and differences of the HI 50.7 data dictionary to the ISO 15926 WIP RDL.
  - *Deliverable 3: Provide written report and proposal for additional work required, based on findings.*

This report is the Task 3 deliverable with updates to information relevant to Task 1 and 2.

2 Background Information

The purposes of this section are (1) to briefly explain the underlying technologies used by AEX and ISO 15926 for the nontechnical reader and (2) to provide more details about the overall design philosophies of AEX, ISO 15926, and the HI-EDE for those unfamiliar with how they relate to each other.

2.1 Brief XML Background

A basic understanding of XML and XML Schema is critical in knowing how the HI-EDE and AEX work together to support interoperability. An XML document is a computer file that contains information in a structured way such that it is easy for computer programs to process, while still possible for a human to understand it as well (as opposed to binary formats). The fundamental unit in an XML document is an *element*, which contains a name, text content, attributes, and possibly child elements that are contained within in it.

Consider a hypothetical world where the only data to be concerned about within the context of a centrifugal pump are flow, head, and the revision string indicating which "version" of the exchanged information is being represented.[3] Below are two examples of how an XML file might represent this information.

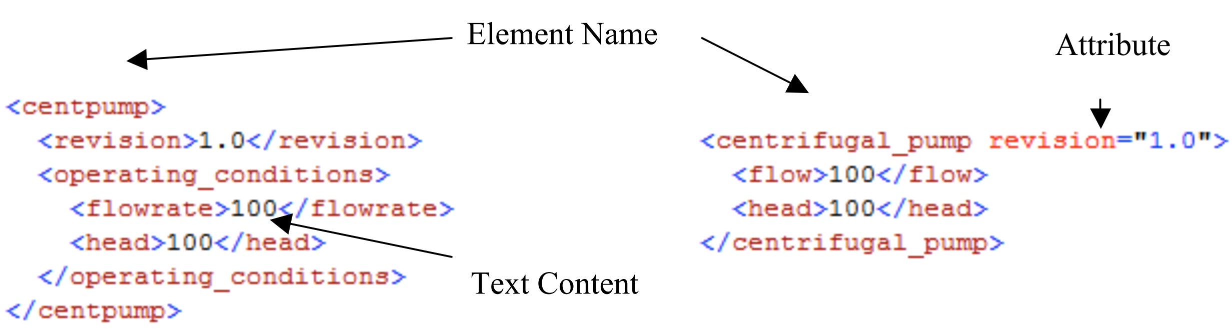

Figure 1: Option A for simple exchange.

Figure 2: Option B for simple exchange.

[3] This example is purposely trivial; real electronic data exchange documents contain far more information and require important design decisions regarding how data should be organized such that it is robust and efficient. This example is meant to illustrate the underlying concept of XML and Schema.

Both options A and B are perfectly reasonable XML representations of the hypothetical dataset. For example, if two companies were to agree to use Option A, their applications would have no trouble exchanging information after completing the appropriate application code to process the structure. The same situation exists for Option B. The difficulty regarding interoperability is that parties involved in the exchange **must** agree on the structure of the XML. If one company produces XML structured as in Option A and the other is expecting Option B, the two companies cannot communicate *even though both of their applications use XML*. While a human reader can plainly see these documents contain the same information (the reader can make the connection that "flowrate" in Option A means the same thing as "flow" in B), a machine **cannot** perform that sort of reasoning.

XML Schema is a language for *specifying* the organization of an XML document[4]. When an XML Schema is written, for example, the structure of the XML is documented and published as a definitive "decision" concerning how to organize the document. The AEX Schema represents such a "decision." It defines the element names, attributes, and structure to use when transacting XML documents containing a wide variety of engineering data. When companies create documents that comply with the AEX Schema, they can be confident that any other company with the ability to work with XML in that form will be able to process the data.

2.2 How does the HI-EDE standard fit in?

If the AEX Schema defines the element names and overall structure of an AEX XML document, why does the HI-EDE standard exist? Two primary answers to this question are:

[4] While XML is used to express the XML Schema language, this is a technical detail and should not confuse the purpose/relationship between an XML Schema document and an *instance* document holding actual engineering data.

1) The AEX schema does not enumerate what information is *required* for a given transaction along the life cycle of the project. Instead, it merely defines how to represent data relevant to a *specific* point in the life cycle. The HI-EDE publishes an industry standard consensus of which data fields are required, desired, or supplemental at the RFQ and Quote stages of a pump's life cycle. This reduces ambiguity in the data exchange process.

2) The AEX Schema is large and encompasses many equipment types. AEX XML documents contain many elements and attributes whose names are similar to each other. While a machine can read AEX XML documents, there can be multiple interpretations of which element contains the data that an application is seeking. For example, there could be several elements in the document containing the word "flow," all containing data related to flow. In order to read the "Rated Flow" of a centrifugal pump from an XML document, an application must know precisely which element contains that data. Elements can be identified within an XML document using an XPath expression. The HI-EDE standard publishes XPath expressions alongside each EDE data field to eliminate any ambiguity.

2.3 Separation of data exchange from data presentation

It is important to understand that the structure and technology representing engineering data during an electronic exchange **in no way limits** the way an *application* can present the data for viewing and manipulation by an end-user. Borrowing from the example above, suppose it is agreed that Option B is the XML structure to be used. Suppose it is also agreed that anytime flow and head are included in an XML document, they are always expressed in the SI base unit set[9] – thus removing any ambiguity of what the true value of the data is.

Now suppose there are two applications, one that presents the data to the user in a datasheet format using U.S. standard units and another application presenting the data in a different format and in SI units. Notice that in the

[9] The HI standard also restricts values appearing in AEX XML documents to always use the SI unit set.

diagram below, the application has mapped the data to the correct field in the user interface and performed the unit conversions as needed. The underlying structure of the data is **decoupled** and nearly irrelevant to the end user. This is a basic and common principle in software engineering.

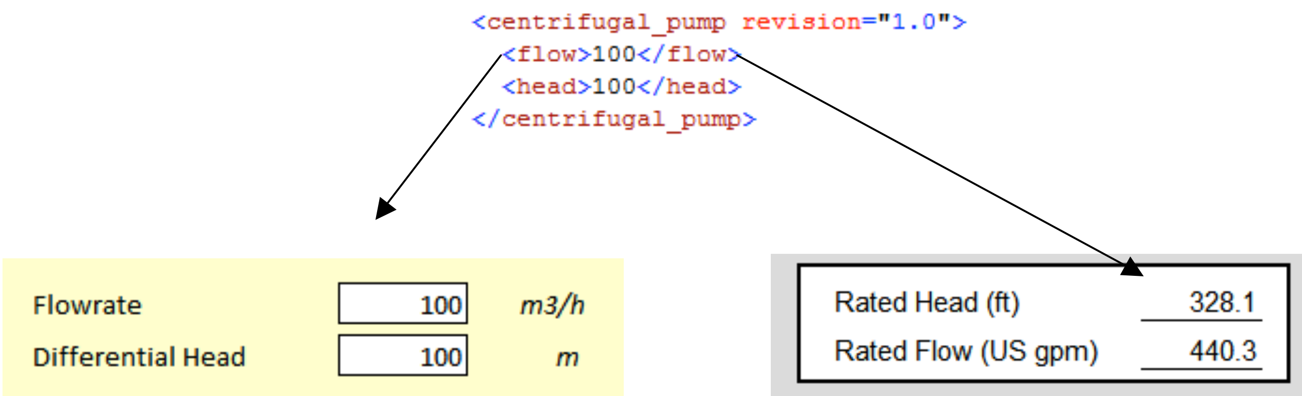

Figure 3: Multiple presentations of the same underlying data.

2.4 Brief Semantic Web Background

When a schema developer creates an XML Schema, he or she *enumerates* the terminology and structure to be used within an XML document. For example, the job of a schema[10] is to codify that the element that contains flow will be named "<flowrate>" and **not** <flow> or <Flow>. Notice, however, that the schema does **not** define what flow actually is (i.e., a measure of the volume of material moving through a system) and the machine processing the XML most certainly has no implicit understanding of this. Furthermore, the application has no innate sense that "head" is in anyway associated with "flow" or any other operating condition. The meaning behind such terms is called "semantics," and "semantic web" is a term to describe a set of technologies aimed at allowing an application to determine the actual semantics behind a specific term through relating it to other terms.

[10] This example is meant to be illustrative – XML schema role and functionality is more significant than portrayed in this simple example.

An “ontology” refers to a set of terms, their definitions, and the relationships between them within a particular domain. An ontology, of course, is data in its own right[11] – and **it** needs to be represented in a computer system. Just as there are several computer storage formats for expressing an xml schema (DTD, XML Schema, RELAX NG), there are several formats for expressing an ontology – with the most popular being OWL/RDF. Coincidentally, both XML Schema and OWL/ RDF are built on XML.

Ontologies are more powerful than simple XML schemas because (1) they do not define the transport mechanism, and (2) data tied to an ontology may be analyzed automatically rather than requiring custom application code to process each individual data field. Once a sufficient set of terms and their relationships to each other have been defined, applications capable of working with the ontology can use the relationship representations to gather additional information about a specific term. For example, relationships might be defined in the ontology that state “flow” is an attribute of “operating condition,” pumps have “operating conditions,” and “head” is also an operating condition. The ability to process the connection between flow and head might allow an application to perform additional checks (i.e., Are all the operating conditions for the pump specified?).

The choice of transportation/exchange format is not tied to the ontology’s implementation. Any storage format that can appropriately relate values to specific terms and relationship is sufficient. Data can be exchanged using spreadsheets, text files, or xml files complying with any schema. **The caveat, however, is that both parties must agree on the specific format and implement code to process it**. **In this way, using semantic web technologies (such as those used in ISO 15926) does not remove the need for schemas to specify the format of a data exchange – it only adds additional contextual information to the underlying data.** Within

[11] An ontology may be thought of as “data that defines data” – which is often called “meta-data”.

the context of the iRING tools, the data format of the exchange is implemented by an "endpoint" – and adapters can be built for text file formats, spreadsheets, and xml documents such as AEX to allow communication between internal and external systems.

2.5 Supporting Electronic Data Exchange

At least two key components of any successful electronic data exchange between two parties are:

1) **(What)** *Deciding what data should be communicated*. This task involves getting domain experts not only to agree on what types of information are pertinent to the exchange, but also what terminology should be used to describe it. As an example, parties must not only agree that "Rated Flow" is an integral piece of data to be exchanged, but also must agree to call it "Rated Flow" and not "Flow, Rated" (or vice versa).

2) **(How)** *Deciding the format and structure in which to communicate such data*. This task is more technical in nature. Is the data to be communicated and stored in a relational database, XML, or some other technology? Once the basic technology is decided, other questions arise, such as what form should the relational database take (tables, columns) or how should the XML Schema be structured?

The AEX project's solutions for "What" and "How" are tightly coupled[12]. AEX Schema defines the structure of the XML document ("How") and also acts as the single governing document for defining what data can be transacted ("What") – including schema annotations and comments for aiding in its interpretation. Parties wishing to communicate must negotiate between themselves which components of the schema they will use.

In contrast, the HI 50.7 EDE fundamentally concerns itself with only the "What" question – for example, defining the "data dictionary" for a centrifugal pump data exchange. The HI-EDE provides a mapping of its

[12] This tight coupling is present in any EDE solution relying entirely on XML Schema.

data dictionary into the AEX Schema – offering a type of "standard" way of interpreting the AEX Schema. The HI-EDE does not provide a technical schema or medium of communication on its own.

ISO 15926 deals with both questions ("What and How"), but does so in a de-coupled way. The Reference Data Library (RDL) contains fundamental classes and templates (which relate classes to each other) that can be built upon to define arbitrary "data dictionaries." Similar to AEX, parties are required to perform some sort of "negotiation" to agree on which segments of the RDL to use. ISO 15926 outlines a variety of methodologies for transacting the data via RDF, SPARQL queries, XML, or even through MS-Excel® spreadsheets.

In many respects, the HI-EDE can be made to augment the ISO 15926 RDL, identifying a standard set of RDL classes and templates for third parties to utilize, much as it identifies a standard mapping to the AEX schema. In addition, the AEX schema could be considered a method of transacting ISO 15926 reference data. In this scenario, AEX XML would be implemented as an endpoint in the iRING Tools system, similar to other endpoints already in existence. The endpoint would be responsible for programmatically transforming RDL entities to and from their AEX counterpart XPaths. The implementation of this transformation is beyond the scope of this initiative.

3 Status of ISO 15926

ISO 15926 is a large, multi-part standard outlining a full life-cycle system for storing, maintaining, and exchanging engineering data at a variety of levels of detail. The first step towards analysis of the standard, along with an initiative to implement it, is identifying the most relevant sources of information. Intelliquip contacted several Fiatech members involved in ISO 15926 at the onset of this initiative to ascertain the most relevant sources. The community was unanimous in its statement that the ISO standards document is not an appropriate starting point. Reasons for this include the depth and verbosity of the standard. This project has been directed towards focusing on the Reference Data Library – the enumeration of "things" that can be represented in ISO

15926 – rather than the transactional technical storage mechanisms such as RDF, OWL, and data facades. The primary sources of information here regarding the concepts behind the data representations are the POSC Caesar Primer and RDS documentation, which are found at the sites below:

- https://www.posccaesar.org/wiki/ISO15926Primer (General Overview of ISO 15926)
- https://www.posccaesar.org/wiki/Rds

Within the scope of this initiative is the assessment of the *differences* between the data represented in ISO 15926 and the HI-EDE (and AEX, through its existing mapping to the HI-EDE). To do this, the ability to search the current Reference Data Library (RDL) is required. Therefore, this project was pointed towards the iRING tools – an open source implementation of ISO 15926. This system, in addition to providing software tools for developing ISO 15926 implementations, provides a reference data editor. The editor, along with documentation about iRING as a whole, is found at the sites below:

- http://iringug.org/wiki/index.php?title=IRINGTools (iRINGTools Homepage)
- http://showroom.iringsandbox.org/Apps/RefDataEditor (iRINGTools Reference Editor)

While we have done preliminary investigations into the transactional and storage mechanisms recommended by ISO 15926 and partially implemented by the iRING tools have been performed during this study, this report solely focuses on the RDL – in that in order to harmonize the HI-EDE with ISO 15926, an alignment of ISO 15926 classes ("what data is represented") is need – not necessarily an implementation of part 8/9/10/11 in the ISO standard.

3.1 ISO 15926 WIP RDL

The HI-EDE defines its data elements at a very high level. For example, a field such as “Best Efficiency Flow Rate for Rated Impeller” carries with it much implicit information – even though the “value” associated with this field is simply a numeric “flow-rate” value. The fact that it is part of the HI-EDE standard first implies that the field represents a part of a centrifugal pump. The field’s definition implies the presence of a physical object (an impeller), along with an efficiency value (not necessarily stated in the transaction) that is achieved at this particular flow-rate. Of course, the impeller itself is not creating the “efficiency” – it is part of a larger system where other components are also playing a role. While the field’s name and value can ultimately be expressed as a simple name/value text pair, the “semantics” behind it are quite complex.

ISO 15926 has an objective of creating a *machine-oriented* methodology for representing this type of data but at a larger scale – encompassing not just centrifugal pumps, but all equipment in a capital facilities plant. ISO 15926 takes a fundamentally different approach in dealing with the semantic complexity of the data, utilizing the concept of machine-readable ontology and semantic web technology to build complex relationships out of low-level “conceptual” building blocks. The ISO 15926 Reference Data Library (RDL) is a distributed repository for specifying these conceptual entities as classes, roles, and templates. Rather than publishing a set of complex fields, ISO 15926 publishes the building blocks – the nouns, verbs, and adjectives that parties may use to build their own field names (and subsequently use the other ISO 15926 technologies to implement the exchange).

A significant difference between the ISO 15926 RDL philosophy and the HI-EDE is that the RDL aims to provide building blocks to communicate all aspects of a plant and ***any*** level of detail – ranging from the procurement agency name down to the size of each screw in each component of each machine. The HI-EDE, conversely, attempts to enumerate specific levels of detail (through the use of required, desired, and supplementary data fields) used in a given transaction (namely the Bid-Quote and Bid-RFQ transactions) concerning a specific slice of the domain (namely pumps). It is this difference that drives the divergent approaches.

3.1.1 Status of RDL

Ultimately, the set of "things" that go into the building of a plant are known – however, they have not necessarily been completely enumerated in a central location. In fact, it may never be possible to develop a complete enumeration of all fields that could be transacted over such a large domain space.

The structure of the RDL embraces "incompleteness," by allowing the RDL to be comprised of a federation of repositories – ranging from industry standard classes to "sandbox" repositories created and possibly discarded within the scope of a single project or transaction. By its nature, the RDL is always "incomplete" in that not every concept is represented. **The assessment in this study suggests that the centrifugal pump domain in the RDL is largely incomplete** – although many concepts that are not specific to centrifugal pumps, but are included in the HI-EDE, are already available *somewhere* in the RDL.

3.1.2 RDL Development Tools

The iRING toolset provides an implementation of ISO 15926. Much of the toolset is dedicated to transacting and storing *instance* data in a variety of formats – centered on Parts 8-11 of the specification. It is this area that the iRING system is most mature.

Another part of the iRING tools focuses on the RDL itself. The Reference Data Editor (below) is a sophisticated web application that allows users to query the entire federated repository of existing classes and templates in the RDL. It also provides editing tools for modifying the RDL. While still under development, the tool appears functional and well-suited for the purposes of the harmonization. It is important to note, however, that although iRING constitutes a well-developed and mature software application for *working with the RDL*, the content of the RDL itself is incomplete. Thus, the work of mapping the HI-EDE into the RDL appears to be a lengthy task.

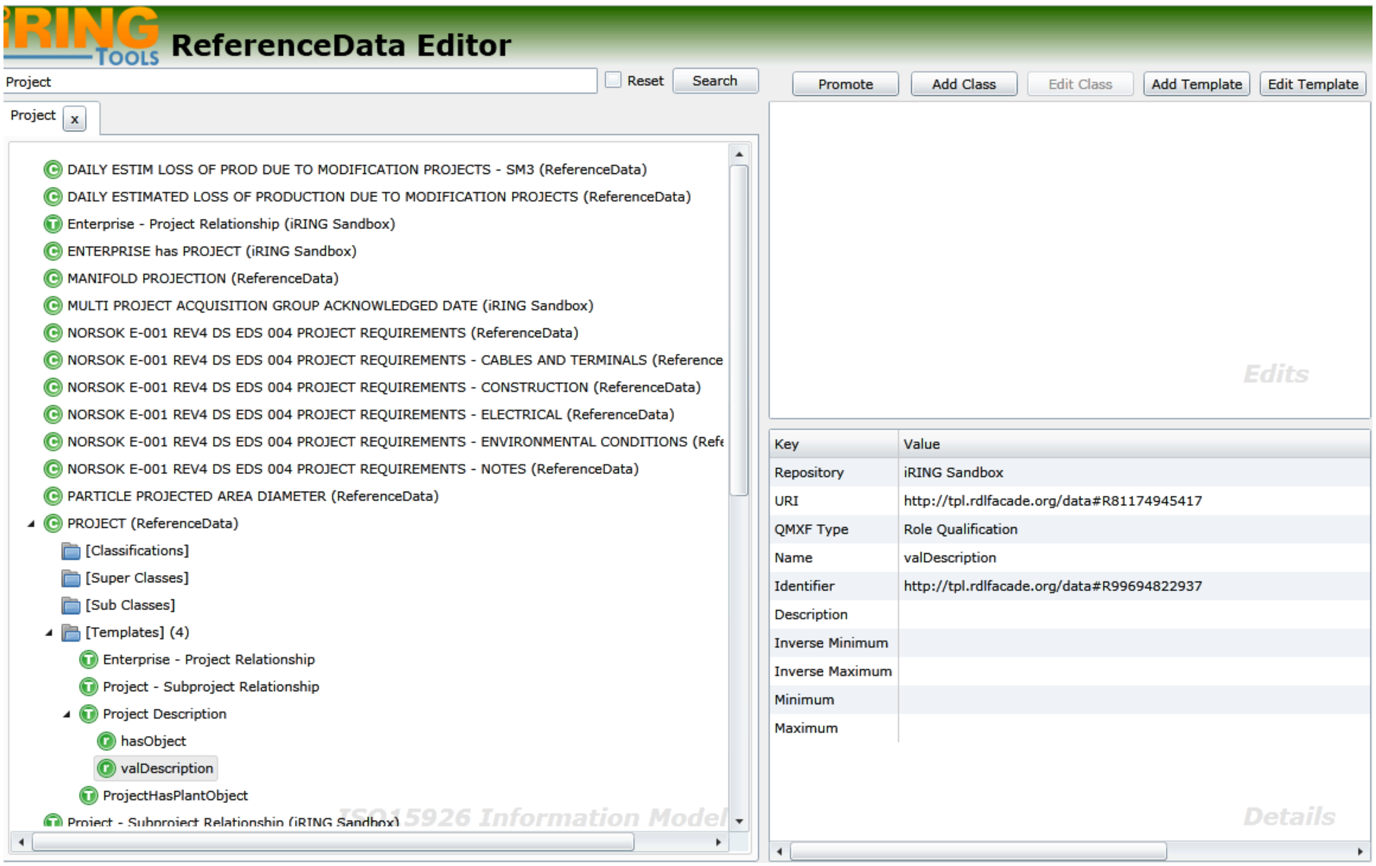

Figure 3: iRING Reference Editor

Through discussions with the iRING development team, the investigators have learned that several feature enhancements are planned, such as the ability to navigate up and down template hierarchies and search for templates relating to subclasses of a given class. These features, among others, will significantly improve the search ability of the RDL and will help during the harmonization process.

4 Aligning AEX, HI-EDE, and ISO 15926

As part of Task 2 of this initiative, Intelliquip organized a series of meetings with Fiatech members involved in the ISO 15926 project. The goal of these meetings was to reach a consensus of how AEX, HI, and ISO 15926 fit together, and most importantly, how the data models can be aligned. This project focused entirely on the process of mapping HI fields to the RDL and did not concentrate on AEX, as each HI field is *already* mapped to the AEX schema. Once each HI field is mapped to the RDL, the mapping between RDL entities and the AEX schema becomes implicit.

An initial meeting was held on April 12, 2011, at the Bechtel facility in Frederick, Maryland. Eighteen representatives from eight different organizations attended. Aside from gaining an initial understanding and introduction to the ISO 15926 standard, attendees focused on mapping a subset of the HI data fields to the RDL. The underlying goal was not to complete a set of mappings. Instead, it was to uncover the complexities involved in the mapping process by forcing all to think concretely about specific fields and specific implementation issues. From the onset of the discussion, it was clear that while the mapping of HI and ISO 15926 appear conceptually straightforward, complexities and design decisions abound. The meeting resulted in a significant amount of information concerning what the process would be and which decisions would need to be made along the way. These efforts led to organizing a more focused group, which met three additional times via web-conferencing (April 26, May 10, and June 15) in order to complete the set of mappings and to record lessons learned. The group consisted of the following members:

Darius Kanga	dkanga@bechtel.com	Bechtel
Robin Benjamins	rxbenjam@bechtel.com	Bechtel
Manoj Dharwadkar	Manoj.Dharwadkar@bentley.com	Bentley
Scott Frees	sfrees@intelliquip.com	Intelliquip
Tryg Dahl	tdahl@intelliquip.com	Intelliquip
Russell Adams	Russell.H.Adams@gmail.com	RHAdams Consulting
Mark Palmer	mark.palmer@nist.gov	NIST

By the conclusion of the sample mappings, the group had reached a consensus on three key questions:

Given and HI-EDE data field…

1) **How would one find a corresponding mapping in the RDL?**

2) **What needs to be done if there is no existing mapping?**

3) **How is such a mapping published in a reusable and accessible manner?**

4.1 How would one find a corresponding mapping in the RDL?

ISO 15926 uses three types of entities to represent concepts: classes, templates, and roles. A class represents a stand-alone entity, such as "Electric Motor." In some situations, a particular HI field will map to a single instance of a class in the RDL. In most cases, however, HI fields contain contextual information that relates several entities together. For example, the field "Electric Motor" does not represent any motor, but represents specifically one driving a centrifugal pump. In ISO 15926, there is a single class representing a centrifugal pump and a template that uses roles to relate the centrifugal pump to the motor. The group's discussions suggest this full expression of the relationship is the best way to express the mapping.

Considering the thousands of classes, templates, and roles in the RDL, it is crucial that a set of standard methodologies for *finding* the mapping be developed. Given the flexibility of the RDL, it is not feasible to deem particular mappings as "best" or "correct" mapping to an HI field – it is more an issue of whether the mapping is appropriate to a domain expert. The group agreed that the best way to locate classes and templates aligned with HI definitions is to use the iRING RDL Reference Data Editor – freely available at http://showroom.iringsandbox.org/Apps/RefDataEditor. This tool allows a user to search the RDL across a number of repositories. In addition, the group determined that the search should be limited to the (1) Proto and Initial, (2) Reference Data, and (3) iRING Sandbox repositories. This ensures the entities used in the HI mapping have been properly vetted by domain experts.

4.2 What needs to be done if there is no existing mapping?

Most HI fields will require classes and templates to be added to the RDL. New entities must be added to the RDL in such a way that they can be made available to the larger community. This study considered several options; however, it was clear that the preferred approach is to create a new sandbox on the iRING site to include new classes. The sandbox is read-only from the web and is edited using an MS-Excel® spreadsheet loading application provided by iRING.

Extending the RDL should be done in concert with as many other groups performing similar tasks. This collaboration is achieved by making sure all RDL projects are federated (i.e., all sandbox RDL sets are accessible to the public via tools such as iRING)[13].

An overriding principle in the RDL model is "specialization." In most situations, rather than adding a completely new entity to the RDL, existing classes and templates "almost" work for the given situation. An example of this may be a template relating Stream to a given Centrifugal Pump. There is no template currently in the RDL that does precisely this. However, a more general template called "AssemblyOfIndividualTemplate" contains two roles – "hasPart" and "hasWhole." Each role can be filled by a class of type "PossibleIndividual."

The "AssemblyOfIndividualTemplate" template is too general to pass an accurate mapping of the relationship between Stream and Centrifugal Pump. Rather than creating a completely new template, however, it is more

[13] There is already a community sandbox at iRINGSandbox.org that could be used to elevate shared RDL items to a broader community level.

effective to *specialize* the template to restrict the role types to be instances of Fluid Stream and Centrifugal Pump classes. The specialization makes finding the new template in the RDL easier and helps the management of future modifications.

Specialization of classes and templates may be executed using the same spreadsheet interface used to create new base classes and templates. The details of the spreadsheet and its associated loading applications have not been included in this document, but are available through the iRING toolset.

4.3 How is such a mapping published in a reusable and accessible manner?

The HI-EDE data dictionary and the corresponding AEX XPaths for each data element are published in Appendix A.1 in the HI-EDE Standard 50.7. Any work undertaken to link the HI-EDE data set to ISO 15926 must be reusable and publishable. Within the ISO 15926 RDL, the only method to identify a class, template, and role uniquely is by its URI[14]. Since an individual HI field's mapping is comprised of a combination of several RDL entities, the most technically accurate way to publish the mapping is by associating a list of URI identifiers with each HI field. While this provides adequate technical information to an implementer, it is decidedly unhelpful for the reader of the publication as it does not convey any information unless the URI values are retrieved using an RDL library, which would provide the names and descriptions.

This study proposes using a hybrid approach, where a human-readable "path" of each RDL class involved in the mapping is enumerated and is accompanied by their URIs and the URI of the templates and roles that create the relationships between the enumerated classes. Below is an example of how field mappings would thus be published:

HI Field: Project/Job Title

[14] http://en.wikipedia.org/wiki/Uniform_Resource_Identifier

The name, or title, of the project/job as assigned by the purchasing organization

ISO 15926 Class Path:			**Pump > Project > Project Description**
[C]	Centrifugal Pump:		http://rdl.rdlfacade.org/data#R84144178538
[T]	Project Has Plant Object:		http://tpl.rdlfacade.org/data#R87985823750
	[R]	Has Part (Pump)	http://tpl.rdlfacade.org/data#R58951773002
[C]	Project:		http://rdl.rdlfacade.org/data#R51988624155
[T]	Project Description		http://tpl.rdlfacade.org/data#R81174945417
	[R]	Has Object: (Project)	http://tpl.rdlfacade.org/data#R32187253595
	[R]	valDescription	http://tpl.rdlfacade.org/data#R99694822937

The mapping begins with what is being called here an "ISO 15926 Class Path." This path represents the classes that will be involved in describing the concept the HI field represents. The class path connects classes with the '>' symbol, which is a placeholder for the suitable template and role combination required to connect the left and right side in the RDL. This class path always starts with the Pump class, as the HI-EDE set deals exclusively with pumps. Note that this representation does not preclude paths created to start with different equipment types and does not mean that equipment-independent data such as "Project Description" would be replicated in an ISO 15926 transaction. This mapping technique is similar to "OIM Path," which was developed and used as part of the iRING Tools project in the past.

While the class path provides a concise notation for orienting the reader in the RDL, a full elaboration of the templates and roles used to connect the classes, along with all of the entity URIs, is required to publish an unambiguous mapping.

The HI 50.7 standard published these RDL mappings. It appears, however, to be less useful for the RDL itself to contain references back to the HI standard[15]. Reasons for this asymmetry include the following.

- The only place to provide this mapping would be in the text description of the classes and templates. The free form nature of this facility makes validation and auditing difficult.
- The RDL is a "living" repository, meaning new classes and templates may be added any time. If multiple classes/templates point themselves to the same HI field, considerable confusion could occur, removing much of the benefit of the mapping.
- It is highly unlikely that the HI will be the only standards organization to map a dataset to the RDL. It is best not to clutter the RDL with references to dozens of external standards.
- Each HI field will map to a combination of classes/templates/role entities in the RDL. This is the nature of the ISO 15926 technology, i.e., concepts are normally represented by a combination of building blocks. While mapping HI to RDL is a one-to-many task, mapping RDL to HI is a many-to-one task and can be a maintenance difficulty.

  The same class/template may appear as part of the mapping in several HI fields. Allowing many HI field designations to appear in the description of RDL entities will be confusing and likely not helpful.

4.4 Full list of HI-EDE to ISO 15926 Data Mapping Samples

A total of eight HI fields, encompassing a broad sampling of the information model, was mapped fully to the ISO 15926 WIP RDL. Below is a schematic diagram of how these fields relate to each other:

[15] It may be possible to enhance the structure and capabilities of the RDL and iRING tools in order to mitigate the maintenance and organization issues a bi-directional mapping would introduce. The enhancements are worth considering, as they could facilitate programmatic use of the mappings at runtime by ISO 15926 compliant systems.

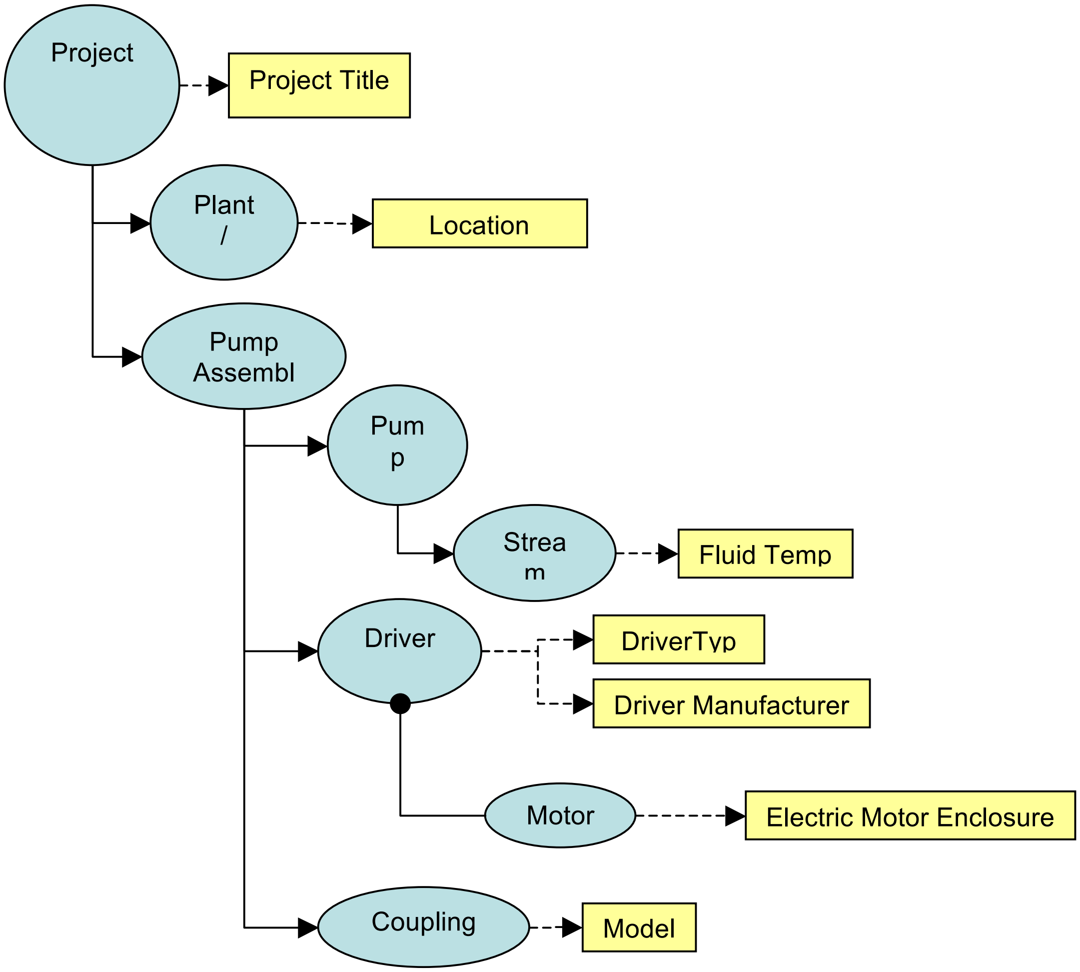

Figure 4: HI Fields used for testing the mapping process to ISO 15926 RDL

In the mappings below, the following symbol notations are used:

[C]	An existing class in the RDL
[C+]	A new class added to the RDL to support HI 50.7
[T]	An existing template in the RDL
[T+]	A new template added to the RDL to support HI 50.7
[R]	An existing role in the RDL
[R+]	A new role added to the RDL to support HI 50.7

Each of the eight HI mappings is listed below in its fully publishable form.

4.4.1 Project/Job Title

HI Field Definition: The name, or title, of the project/job as assigned by the purchasing organization

ISO 15926 Class Path: **Pump > Project > Project Description**

[C] Centrifugal Pump: http://rdl.rdlfacade.org/data#R84144178538

[T] Project Has Plant Object: http://tpl.rdlfacade.org/data#R87985823750

[R] Has Part (Pump) http://tpl.rdlfacade.org/data#R58951773002

[C] Project: http://rdl.rdlfacade.org/data#R51988624155

[T] Project Description http://tpl.rdlfacade.org/data#R81174945417

[R] Has Object: (Project) http://tpl.rdlfacade.org/data#R32187253595

[R] valDescription http://tpl.rdlfacade.org/data#R99694822937

4.4.2 Pumped Fluid Temp, Max

HI Field Definition: The maximum fluid temperature to which the equipment described herein can be exposed. The temperature is to be provided by the owner.

ISO 15926 Class Path: **Pump > Stream > Maximum Temperature**

[C] Centrifugal Pump: http://rdl.rdlfacade.org/data#R84144178538

[T+] CentrifugalPumpStream: http://tpl.rdlfacade.org/data#R29983017220

[R] hasPart (FLUID STREAM) http://rdl.rdlfacade.org/data#R35385587762

[R] hasWhole (CENTRIFUGAL PUMP) http://rdl.rdlfacade.org/data#R84144178538

[C] Stream http://rdl.rdlfacade.org/data#R94736743852

[T+] MaximumTemperatureStream: http://tpl.rdlfacade.org/data#R29024048968

[R] hasPossessor (FLUID STREAM) http://rdl.rdlfacade.org/data#R35385587762

[R] hasType (MAXIMUM TEMPERATURE) http://tpl.rdlfacade.org/data#R56806886394

[R] valValue (double) http://www.w3.org/2001/XMLSchema#double

[R] hasScale (TEMPERATURE SCALE) http://rdl.rdlfacade.org/data#R59174773370

[C] Maximum Temperature: http://rdl.rdlfacade.org/data#R80389456724

4.4.3 Site Location

HI Field Definition: The name of the geographic area where the site is located

ISO 15926 Class Path: Pump > Plan Location

[C] Centrifugal Pump: http://rdl.rdlfacade.org/data#R84144178538

[T] Plant Location has Plant Object: http://tpl.rdlfacade.org/data#R56493865993

[R] hasPart: http://tpl.rdlfacade.org/data#R58951773002 (Pump)

[R] hasWhole http://tpl.rdlfacade.org/data#R85898998707 (Plant Site)

[C] Plant Location: http://rdl.rdlfacade.org/data#R74863378857

4.4.4 Driver Type

HI Field Definition: A choice of the driver type as assigned by the owner. If blank, the supplier should indicate the type of driver that will be used.

ISO 15926 Class Path: Pump > Pump Driver > DriverType

[C] Centrifugal Pump: http://rdl.rdlfacade.org/data#R84144178538

[C] Pump Driver: http://rdl.rdlfacade.org/data#R21302049621

[T+] PumpToDriverOfTypeConnection: http://tpl.rdlfacade.org/data#R68350050141

[R] hasConnectionType: http://rdl.rdlfacade.org/data#R20735180747 *

[R] hasSide1: http://tpl.rdlfacade.org/data#R62330117117 (PUMP)

[R] hasSide2: http://tpl.rdlfacade.org/data#R78459697944 (PUMP DRIVER)

** The connection type role for this template has been restricted to drivers of classification PUMP DRIVER TYPE CLASS since this is an enumeration within the HI-EDE Standard. This new classification has been created in the RDL. Each choice in the enumeration maps to existing classes within the RDL, as shown below:

Members of PUMP DRIVER TYPE CLASS (Class of Class): URI: http://rdl.rdlfacade.org/data#R89907450825

Diesel Engine:	URI	http://rdl.rdlfacade.org/data#R87348972627
Electric Motor:	URI	http://rdl.rdlfacade.org/data#R16224749002
Engine:	URI	http://rdl.rdlfacade.org/data#R20755995734
Gas Engine:	URI	http://rdl.rdlfacade.org/data#R45940272580
Expander:	URI	http://rdl.rdlfacade.org/data#R62667926700
Gas Turbine:	URI	http://rdl.rdlfacade.org/data#R99456935879
Hydraulic Turbine:	URI	http://rdl.rdlfacade.org/data#R14594265647
Hydraulic Motor:	URI	http://rdl.rdlfacade.org/data#R52815362952
Steam Turbine	URI	http://rdl.rdlfacade.org/data#R49484528110
Turbine:	URI	http://rdl.rdlfacade.org/data#R21209598935

4.4.5 Driver Manufacturer

HI Field Definition: Driver Manufacturer company name. If blank, the supplier should indicate the driver manufacturer that will be used.

ISO 15926 Class Path:		**Pump > Pump Driver > Manufacturer > Manufacturer Name**
[C]	Centrifugal Pump:	http://rdl.rdlfacade.org/data#R84144178538
[T]	Pump to Driver Connection:	http://tpl.rdlfacade.org/data#R85736598359
[C]	Pump Driver:	http://rdl.rdlfacade.org/data#R21302049621
[T+]	Pump Manufacturer Relation	http://tpl.rdlfacade.org/data#R48899402955

[R+] hasObject: http://rdl.rdlfacade.org/data#R20735180747

[R+] hasFabrication http://rdl.rdlfacade.org/data#R13646688670

[C] Manufacturer http://rdl.rdlfacade.org/data#R13646688670

[T] Manufacturer Name: http://tpl.rdlfacade.org/data#R57455493984

[R] hasObject: http://tpl.rdlfacade.org/data#R57455493984

[R] valIdentifier: http://tpl.rdlfacade.org/data#R22674749688 (Manufacturer Name)

4.4.6 Coupling Model, motor to pump

HI Field Definition: Coupling model, using the coupling manufacturer's nomenclature.

ISO 15926 Class Path: **Pump > Driver > Coupling**

[C] Centrifugal Pump: http://rdl.rdlfacade.org/data#R84144178538

[C] Coupling http://rdl.rdlfacade.org/data#R78254885518

[C] Pump Driver: http://rdl.rdlfacade.org/data#R21302049621

[T+] DriverCoupling http://tpl.rdlfacade.org/data#R59022718551

[R+] hasSide1 (PUMP): http://rdl.rdlfacade.org/data#R21302049621

[R+] hasSide2 (Coupling): http://rdl.rdlfacade.org/data#R78254885518

4.4.7 Gear Box Size

HI Field Definition: Size designation of the gearbox using the manufacturer's standard sizing nomenclature.

ISO 15926 Class Path: **Pump > Gearbox > Size**

[C] Centrifugal Pump: http://rdl.rdlfacade.org/data#R84144178538

[C] GEAR BOX: http://rdl.rdlfacade.org/data#R82669801563

[T+] CentrifugalPumpHasGearbox http://tpl.rdlfacade.org/data#R31537619028

[R] hasWhole (Centrifugal Pump) http://rdl.rdlfacade.org/data#R84144178538

[R] hasPart (GEAR BOX) http://rdl.rdlfacade.org/data#R82669801563

[T+] GearboxSize http://tpl.rdlfacade.org/data#R67782548677

[R] hasPossesor (Gearbox) http://rdl.rdlfacade.org/data#R82669801563

[R] hasProperty (MODEL AND SIZE) http://rdl.rdlfacade.org/data#R92145430548

4.4.8 Electric Motor Enclosure

HI Field Definition: The classification according to environmental protection and methods of cooling.

ISO 15926 Class Path: **Pump > Motor Enclosure**

[C] Centrifugal Pump: http://rdl.rdlfacade.org/data#R84144178538

[C] MOTOR ENCLOSURE http://rdl.rdlfacade.org/data#R92043504852

[T+] CentrifugalPumpMotorEnclosure http://tpl.rdlfacade.org/data#R21983502743

[R] hasWhole (Centrifugal Pump) http://rdl.rdlfacade.org/data#R84144178538

[R] hasPart (MOTOR ENCLOSURE) http://rdl.rdlfacade.org/data#R92043504852

4.5 Working with Picklists

For any engineer used to working with datasheets, the use of checkboxes on a paper datasheet is quite familiar. These provide a set of valid choices to the person filling out the information. Electronic datasheets employ similar techniques to limit choices, but more user interface options are available (radio buttons, checkboxes, picklists/dropdowns). The advantage over using plain text boxes is clear – these systems allow for a much more standardized set of values to appear on the datasheet, helping to limit ambiguity.

The same principle is applicable in electronic data exchange. The HI-EDE specifies enumerations of textual values for many HI-EDE data fields. Typically, datasheets represent these fields using radio buttons, checkboxes, and picklists. However, that is not necessarily a requirement. The key requirement is that when *exchanging* these fields, only the values enumerated in the HI-EDE standard can be transmitted[16].

Ultimately, the underlying transport technology needs to express these restrictions. When working with AEX, the schema itself restricts corresponding XML elements in order to support the enumerations. The ISO 15926 RDL supports this restriction in a vastly different way. In the RDL, *each choice* is not simply represented by a string value. Each possible value is its own RDL class. The RDL supports a concept called classifications, which is a flexible method of expressing that a collection of classes share *something* in common. This is **not** an inheritance or polymorphic relationship – it is merely a categorization. Classes can belong to many different classifications. Wherever a particular value must be restricted to a set of choices, the class type for the template role representing the value is set to the category, not a specific class. The term commonly used for this is "class of class."

The HI field "Driver Type" is a classic example of an enumeration. The HI standard defines several types of drivers – Electric Motor, Gas Engine, Gas Turbine, and others. A specific XML element in AEX stores driver type and the schema places a restriction on it such that it can *only* be of the values specified in the standard. In ISO 15926, individual classes are in the RDL for each option (see the full listing above). There is also a "class of classes" called PUMP DRIVER CLASS. Classes such as Electric Motor, Gas Engine, and others are assigned the PUMP DRIVER CLASS classification such that the set of all classes in that classification align with the valid choices set forth by the HI standard. Finally, any template that uses "Driver Type" as one of its roles defines the type of such role to be PUMP DRIVER CLASS, not a specific class. At runtime, any class with the PUMP DRIVER CLASS classification can fulfill that role.

---

[16] The choice of "Other" is often listed to allow a free-form, non-standard value to be placed in the field – however its use is not encouraged.

5 Conclusions

5.1 Achieving Interoperability with ISO 15926 RDL

When working with ISO 15926, the parties involved in the data exchange **must mutually agree on how they will use the RDL to represent their information**. The "negotiation" is *not automated* and is embraced as part of the design philosophy of the standard. While less pronounced, a similar negotiation is required when using the cfiXML AEX schema. To transact data, trading partners must make sure they agree on a *bounded set of data fields* to transact and *where* each can be represented in the schema (i.e., they must agree on an XPath for the data they are exchanging).

The HI recognized that while this negotiation step cannot be avoided for complex use cases, it is valuable to have an "industry standard" bounded set of high value fields overlaid on the AEX schema so that companies wishing to begin implementing electronic data exchange can have confidence that their implementations will be compliant with another's systems. Thus, the HI published its data set with thoroughly vetted XPath mappings into AEX. While each mapping may not be the only possible interpretation, the fact that "one" valid mapping has been published and accepted by the community allows unambiguous data transfers to be made more easily.

The study here envisages the HI 50.7 standard playing a similar role for ISO 15926. Like AEX, ISO 15926 covers a far more broad information model than HI (where the focus is on pumps). Both AEX and ISO 15926 provide enormous flexibility but can be daunting for initial implementers - making mapping the HI 50.7 dataset to the ISO 15926 RDL critically valuable. Just as with AEX, mappings into ISO 15926 RDL made by the HI may not be the only options for some data fields, but by offering a published "industry standard,"confidence can be instilled in early adopters of the ISO 15926 ecosystem that pump data can be exchanged with trading partners unambiguously. In addition, the information model in the ISO 15926 RDL is missing much of the data elements required to cover the HI 50.7 standard, and the mapping process would force the RDL to become more complete.

A mapping of an HI field into the ISO 15926 WIP RDL is not a one-to-one mapping. An HI field represents a rather complex concept, with a great deal of engineering context behind it. To represent this in the RDL, one must combine sequences of classes and templates. The level of alignment between the HI 50.7 and ISO 15926 standards is more accurately recast with the question: "Does the ISO 15926 RDL have the building blocks to express HI 50.7?" The answer to this is "somewhat." Many of the building blocks are indeed present, but many also are missing.

While the HI exclusively deals with pumps, other standards organizations have a role to develop similar standard data dictionaries for other equipment types. These organizations may also map to AEX and the ISO 15926 standard and work together to use the same mappings for common "equipment-neutral" data such as "Project Name" and "Site Location." One of the next logical equipment types to investigate is likely motor operated valves, as previous Fiatech initiatives have already begun the process of creating a data dictionary for valves similar to HI's work with pumps. This is seen as a highly valuable next step as it will allow the assessment of how multiple equipment types can share overlapping segments of their bounded datasets.

5.2 Defining an HI / AEX / ISO 15926 Harmonization

Based on this research and extensive discussions with those engaged in ISO 15926, at the following reasonable answers are offered to the following important questions:

1) What does an HI-EDE -> ISO 15926 mapping look like?

The HI-EDE can publish its mapping into the RDL. A *set* of URI identifiers will uniquely identify classes, templates, and roles. The publication of these URIs along with a brief textual explanation provides similar information to an implementer as providing XPaths into the AEX Schema.

2) How does this study view a mapping within the RDL?

The iRING Reference Data Editor appears to be the most logical choice of tools for finding concepts in the RDL. Although the editor allows for searching across all federated repositories, the HI mapping shall limit itself to the (1) Proto and Initial, (2) Reference Data, and (3) iRING Sandbox repositories such that it can be sure it is working with properly vetted reference data.

3) How are new entities added to the RDL?

The most logical mechanism for adding new entities required for HI-EDE mappings is to create a separate "HI-EDE Sandbox," which would be included in the federation available via iRING tools to house the additional reference data. iRING servers are a good candidate for hosting the sandbox. It is important to reuse and build upon existing reference data through class and template specialization wherever possible in order to maximize the integrity and impact of the additions.

5.3 Assessment of remaining harmonization work

Completing the harmonization of the HI-EDE, AEX, and ISO 15926 RDL requires a mapping of the entire HI-EDE data set into existing and new entity combinations within the ISO 15926 RDL. It also requires the accurate and accessible publication of such mappings in an electronic form. This research found proper mapping of an individual data field takes an average of 30 minutes, although the time for each field varies tremendously.

It is important to note that providing a mapping of the HI-EDE data fields into the RDL does not provide an immediately usable system. The ISO 15926 standard defines an RDL. However, systems still must be implemented to exchange the data. This requires implementation of appropriate technologies such as XML, RDF, SPARQL, and OWL and is not included in this initiative. The same can be said of the adoption of AEX as well: simply having the mappings of HI fields to AEX XPaths does not complete the implementation of data exchange. Significant software development is required in order to process the XML documents accordingly. Using the iRING (.NET platform) tools partially mitigates implementation effort, but only *after the mapping to the RDL* is complete.

Fiatech

3925 West Braker Lane (R4500)
Austin, TX 78759-5316

t: (512) 232-9600

www.fiatech.org

Construction Industry Institute®
The Cockrell School of Engineering
The University of Texas at Austin