CABA™ Continental Automated Buildings Association

Project-Haystack

A CABA WHITE PAPER

Dan Prairie
Nexus E3 Systems

John Petze
Project-Haystack

Marc Petock
Project-Haystack

Mathan Kumar Gopal Samy
Philips

CABA Intelligent & Integrated Buildings Council

Project Haystack

Connect to what's next™

www.caba.org

Continental Automated Buildings Association

Project-Haystack
A CABA White Paper

Authors

Dan Prairie
Nexus E3 Systems

John Petze
Project-Haystack

Marc Petock
Project-Haystack

Mathan Kumar Gopal Samy
Philips

Working Group

Bernhard Isler
Siemens Industry, Inc.

Dan Prairie
Nexus E3 Systems

Greg Walker
CABA

John Petze
Project-Haystack

Ken Wacks
Ken Wacks Associates

Marc Petock
Project-Haystack

Marek Dziedzic
Public Works & Government Services Canada (PWGSC)

Mathan Kumar Gopal Samy
Philips

Sub-Committee

Ken Wacks (Chair)
Ken Wacks Associates

Brittany Hack
Independent

David Katz
Sustainable Resources Management Inc.

Dilip Sarangan
Frost & Sullivan

Marek Dziedzic
Public Works & Government Services Canada (PWGSC)

Raphael Imhof
Siemens Industry, Inc.

Working Group: Individuals who either contributed ideas and input into the direction of paper or reviewed the final draft.
Sub-Committee: Under the direction of the Sub-Committee Chair, this formal committee reviewed and approved both the initial white paper proposal and final draft.

CABA Intelligent & Integrated Buildings Council

Published: February 2016

CABA Research Program

ABOUT CABA

The Continental Automated Buildings Association (CABA) is an international not-for-profit industry association, founded in 1988, and dedicated to the advancement of intelligent home and intelligent building technologies. The organization is supported by an international membership of over 330 organizations involved in the design, manufacture, installation and retailing of products relating to "Internet of Things, M2M, home automation and intelligent buildings". Public organizations, including utilities and government are also members. CABA's mandate includes providing its members with networking and market research opportunities. CABA also encourages the development of industry standards and protocols, and leads cross-industry initiatives. CABA's collaborative research scope evolved and expanded into the CABA Research Program, which is directed by the CABA Board of Directors. The CABA Research Program's scope includes white papers and multi-client market research in both the Intelligent Buildings and Connected Home sectors. www.caba.org.

ABOUT CABA'S INTELLIGENT & INTEGRATED BUILDINGS COUNCIL (IIBC)

The CABA Intelligent & Integrated Buildings Council works to strengthen the large building automation industry through innovative technology-driven research projects. The Council was established in 2001 by CABA to specifically review opportunities, take strategic action and monitor initiatives that relate to integrated systems and automation in the large building sector. The Council's projects promote the next generation of intelligent building technologies and incorporate a holistic approach that optimizes building performance and savings. www.caba.org/iibc.

DISCLAIMER

This white paper was developed and published by CABA for the industry with permission from the authors. CABA expresses its appreciation to the authors and contributors for making this white paper available to be included as part of CABA's Members Library and CABA's Public Library. CABA, nor any other person acting on their behalf of CABA assumes any liability with respect to: the use of, or for damages resulting from the use of, any information, equipment, product, method or process disclosed in this white paper.

This CABA White Paper and other industry research reports can be found in CABA's Members Library and CABA's Public Library at: http://www.caba.org. This information is also keyword searchable. Contact the CABA office if you do not have the passwords to access this material by email caba@caba.org or phone 888.798.CABA (2222) or 613.686.1814 x228. CABA encourages you to share this white paper with others in your organization and the industry. Permission is not required from CABA to share this white paper, as long as proper acknowledgment is provided to CABA.

PUBLISHED

March 2016.

TABLE OF CONTENTS

Published: February 2016

1. INTRODUCTION

The Internet of Things (IoT) is changing our world, including affecting the way we manage and operate our buildings. Lower cost sensors, more powerful controllers, the "Cloud", smart equipment, and new software applications are enabling new approaches to facility management. Smart devices are creating dramatic increases in the amount and type of data available from our facilities, while new software applications are creating new ways to benefit from that data. Together these advances are driving a fundamental shift in how we manage and operate buildings, enabling us to move from traditional control strategies based on simple feedback loops to data-driven methodologies that inform operators and service providers about the actual performance of building systems in real time. All of these trends can help improve efficiency and drive savings in both energy and overall operational costs when used together effectively.

It's one thing to have access to data; it's another to make it actionable. With more data available than ever before the industry is presented with a new challenge. Device data are stored and communicated in many different formats. It has inconsistent, non-standard naming conventions, and provides very limited descriptors to enable us to understand its meaning. Simply put, the operational data from smart devices and equipment systems lacks information to describe its own meaning. Without meaning, a time consuming manual effort is required before that data can be used effectively to generate value. The result is that the data from today's devices, while technically "available", is hard to use, thus limiting the ability for building operators to fully benefit from the value contained in the data.

1.1. Describing the Meaning of Data – Data Semantics

In order to utilize operational data in value-added applications such as analytics, analysis, visualization and reporting tools, we need to know the meaning of our data. For example, if we obtain a data item from a building automation system (BAS) and it has a value of 77.6 we can't do any effective analysis until we understand whether it is 77.6 degrees F, or degrees C, or PSI, RPM, or kW, etc. "Units" therefore is one good example of an essential descriptor we need in order to understand and use our data, but it is by no means the only one.

Continuing with our example, if all we know are the units (Degrees F for example), we still don't know much about the significance of the value 77.6. If it's a zone temperature it might be a bit warm for occupants. If it's a return air temperature it's right where we want it to be.

Let's say the sensor with the value of 77.6 is named **zn3-wwfl4**. If I am intimately familiar with the building system and the naming conventions used when it was installed I may be able to determine that means Zone 3, West Wing, Floor 4. That would give me a bit of information to work with. If I know the building really well I may also be able to tell that zn3-wwfl4:

- Is a zone temperature
- Is an exterior zone
- Is south facing
- Is supplied by a VAV box

- Is served by AHU-1
- Is operated on occupancy schedule #1, which is 7:30 AM - 6:30 PM
- Has an occupied cooling setpoint of 74 degrees F

Armed with this additional information I could determine that a value of 77.6 is not proper for 9:00 AM on a weekday - it's too hot and will lead to occupant complaints. What enabled me to make that determination, however, was a significant amount of information about the meaning of the specific sensor. Information that I happened to have because of my personal knowledge of the building - but information that was not recorded in the control system (or any single location), and is not available in any consistent "machine readable" format. Herein is the challenge to using the wealth of data produced by today's systems and devices - the ability to represent, communicate and interpret the meaning of data. This "data about data" is often referred to metadata or data semantics

Having appropriate metadata about the sensor point zn3-wwfl4 would enable us to understand the impact of the current value of 77.6 without relying on personal knowledge of the building. As described above, if we knew the schedule associated with the room that the sensor is associated with, we could determine that it is over temperature during occupied hours and the occupant is most likely uncomfortable.

Without the necessary metadata, however, we can't determine the impact of the current value and its relationship to proper system operation. So in order to provide effective use of sensor data we need to combine metadata with the sensor value zn3-wwfl4. When done manually, this process is referred to as mapping or "data wrangling". This step in the utilization of machine data has historically been a time consuming manual process that adds significant cost to the implementation of new software applications such as analytics and data visualization.

Interestingly, with all of the power they have gained over the last decade, and the adoption of standard communication protocols, most building automation systems provide little to no ability to capture semantic information about the data they contain. There has been no standardized approach to representing the meaning of the data they generate or contain. The systems provide points with a name, which typically is ad hoc and follows no universal standard, a value, and units but little other information. The result is that a labor intensive process is required to "map" the data before any effective use of the data can begin. Clearly, this creates a significant barrier to effective use of the data available from smart devices.

1.2. A Metadata Example

So how can we capture all of this information, share it and associate it with the data items in our automation systems and smart devices? We cannot do it simply by trying to use standardized point names. Even in our simple example, we have more metadata that can be effectively captured in a point name. Add to that the fact that we may want to add numerous other metadata items over time that go far beyond our simple example and it's obvious we need another approach. An effective solution needs to have the following characteristics:

1. It should de-couple the point name from the representation of the associated metadata. The concept of tags to represent the metadata works well here. Tags represent "facts" about data items and can be associated with the point names to provide the semantics that describe the point. They tell us about the meaning of the point, but they do not replace the point name or change it in any way. This is essential for any solution to work with existing systems. The reality is that we have millions of points in thousands of systems and their point names cannot be changed. It's simply not an option – and it isn't necessary. What is needed is a standardized model for associate metadata with those existing data items to enable us to associate meaning with the existing point names.
2. It should utilize a standardized library of tags to provide consistency of metadata terminology. This will enable automated tools to interpret data meaning. The library needs to be able to be updated by industry experts as new applications are encountered. The metadata methodology therefore needs to be extensible.

Given our earlier example, a record representing the point **zn3-wwfl4** with its associated metadata might look like the following:

id:	150a3c6e-bef0ee0e	(RecId)
dis:	zn3-wwfl4	(Str)
sensor:	✓	(Marker)
air:	✓	(Marker)
temp:	✓	(Marker)
unit:	°F	(Str)
curVal:	77.60 °F	(Number)
equipRef:	Carytown AHU-4	(Ref)
siteRef:	Carytown	(Ref)
tz:	New_York	(Str)
zone:	✓	(Marker)
vav:	✓	(Marker)
floor:	4	(Number)
scheduleRef:	occSchedule-1	(Ref)

Figure 1 Example of a record representing the point zn3-wwfl4 with associated metadata

The "point" has a unique ID and then is described by a series of tags. In our example, we have the following:

- **dis**: The display name we want associated with the sensor point. This could be the original descriptor from the system supplying the data, or it could be a new name/descriptor.
- **sensor**: a marker tag that tells us this is a sensor point (versus a control point).
- **air**: a tag that tells us this sensor is measuring air.
- **temp**: a tag telling us this sensor is measuring temperature (versus pressure for example.
- **unit**: a tag that tells us the units of measure.
- **curVal**: a tag that holds the most current value supplied by the system.

- **equipRef**: a tag that associates this point with a specific piece of equipment.
- **siteRef**: a tag that tells us the site (building) this point is associated with.
- **tz**: a tag that tells us the time zone of the site.
- **zone**: a tag that tells us this is a zone (space) in a building.
- **vav**: a tag that tells us that this sensor is measuring a zone supplied by a VAV box.
- **floor**: a tag that tells us which floor of the building the sensor is located.
- **scheduleRef**: a tag that tells us which occupancy schedule is associated with the area this sensor is measuring.

The indicators in parenthesis tell us the type of tag: recID: RecID - a unique identifier, Number, String, Marker, and Reference, which points to an associated entity.

With these tags it is now possible for both humans and machines to interpret the meaning of this point. This enables streamlined value creation. For example, a graphical visualization software application could interpret this information to automatically generate a graphical view of the sensor within the appropriate zone. An analytics application could automatically apply appropriate rules to this sensor to identify improper conditions.

1.3. The Role of Project-Haystack in Addressing the Metadata Challenge

The role of Project-Haystack is to bring together a community of industry constituents in a collaborative effort to develop a standardized approach to representing and using metadata across a wide range of applications. Project-Haystack standardizes semantic data models and Web services to access and communicate those models, with the goal of making it easier to unlock value from the vast quantity of data being generated by the smart devices and equipment systems that permeate our homes, buildings, factories, and cities. Applications include automation, control, energy, HVAC, lighting, and other environmental systems.

The Project-Haystack initiative was started in 2011 to directly address this need. It is structured as an open source, community-driven initiative modeled on the open source specifications found in the software and IT industry.

The Project-Haystack vision is to streamline the use of data from the Internet of Things including, but not limited to, building and energy systems, by creating a standardized approach to defining "data semantics" and related services and APIs to consume and share the data and its semantic descriptors. Project-Haystack makes it easier to unlock value from the vast quantity of data being generated by smart devices by making data "self-describing".

With a common methodology for defining metadata, and a common vocabulary or taxonomy (libraries of community defined tags) to give meaning to the data collected by smart devices and systems, we can more easily take advantage of a wide range of software tools to visualize, analyze and generate value from operational data.

1.4. An Open-Source, Community-Driven Approach

Project-Haystack is operated as an open source project, which makes it easy for anyone to get involved. Anyone can easily take advantage of the work of Project-Haystack and contribute to it. All collaboration is done on the discussion forum at http://project-haystack.org. Anyone can contribute on the forum by signing up on the Web site. Domain experts in a given space such as chillers, data centers, or refrigeration can join or start a discussion. Equipment manufacturers who would like to see specific tag models for their products are also a great source of input. All of the work done by Project-Haystack is easily available to the industry community. It can be downloaded without even registering an account on the Web site. There is no cost or obligation associated with using Project-Haystack techniques, tagging libraries and open source reference implementations.

1.5. Project-Haystack is More Than Meta Data Tagging

It's important to note that Haystack is more than one thing. First, it's the data modeling methodology - the simple, flexible tagging approach that can be used in media from Excel spreadsheets and CSV text files, to data tables in embedded devices, XML representations, Web services and others.

Second, it's the consensus developed tagging libraries (taxonomies) published and made available for download and use (at no cost). You can find all of the tagging libraries developed by the Project-Haystack community here: http://project-haystack.org/tag.

Third, it's the REST[1] communication protocol designed to exchange Haystack tags between applications.

Fourth, it's the software reference implementations and complementary applications being developed by various community members and companies. As of the date of publication these included:

- **Haystack Java Toolkit**: lightweight J2ME compliant client and server implementation
- **NHaystack**: Niagara module to add Haystack tagging and the Haystack REST API
- **Haystack CPP**: C++ Haystack client and server implementation
- **Haystack Dart**: client library for Dart programming language
- **NodeHaystack**: node.js client/server implementation

You can find links to download all of these software reference implementations here: http://project-haystack.org/download.

Perhaps most important, however, is the community that has formed to address the challenges of data modeling for building systems and IoT devices. The Project-Haystack community continues to grow and expand the equipment and device models (taxonomies) and extend the range of applications served by Project-Haystack

[1] Representational state transfer (REST) is the software architectural style of the World Wide Web.[1][2][3] REST's coordinated set of constraints, applied to the design of components in a distributed hypermedia system, can lead to a higher-performing and more maintainable software architecture. To the extent that systems conform to the constraints of REST they can be called RESTful. RESTful systems typically, but not always, communicate over Hypertext Transfer Protocol (HTTP) with the same HTTP verbs (GET, POST, PUT, DELETE, etc.) that Web browsers use to retrieve Web pages and to send data to remote servers.[4] REST systems interface with external systems as Web resources identified by Uniform Resource Identifiers (URIs). Source: Wikipedia.

A Flexible Extensible Model that Can Be Used Beyond Consensus Approved Tags. It is also important to mention that the Project-Haystack methodology can be used beyond just the consensus approved tags available at any point in time. Projects and products can add custom tags to represent information important to their needs without requiring submission or approval. And those custom tags can still be discoverable and interpretable by other applications with minimal effort to add a reference index or look up table. Applications that are "Haystack tag-aware" can be easily extended to interpret new community approved tags and custom tags as they are added. Numerous companies have started their adoption of Haystack by developing their own haystack compatible tagging libraries and then returned that work to the community over time.

1.6. What Will the World Look Like When We Have Standard Tagging Models to Describe Building, Equipment and Device Data?

Today, analyzing building system data requires significant time and effort because these contextual models must be built by hand. This means that the opportunity to use data to reduce energy use, and improve operational efficiency and performance is being stifled by the labor costs associated with manual mapping of system data and the inconsistency of those efforts. Project-Haystack's mission is to define this common vocabulary so that we can build models of our buildings and systems, enabling us to more efficiently derive value from all the data our building automation systems and smart devices are producing. With Project-Haystack models we can transition from a manual process to an automated process and unlock the value of our operational data.

With metadata "tagging" a wide variety of software applications can now automatically find and interpret the data they need to provide value to the user. Two specific examples of the benefits of metadata tagging being demonstrated today include:

- Software today can automatically generate equipment graphics and system views simply by interpreting tags on the control system data. Hours of manual graphics assembly is thereby eliminated reducing project cost and increasing value creation.
- With proper metadata, analytics applications can quickly interpret patterns in operational data to identify faults, deviations, and trends that can be addressed to improve efficiency and insure proper operation of equipment systems.

Project-Haystack enables a future where a push of a button can turn data into true intelligence, reducing the cost of intelligent building systems and enabling operational teams to better understand and improve the operation and performance of buildings.

2. THE IMPORTANCE OF DATA - HOW DATA IS CHANGING THE BUILDING INDUSTRY

Data has become the oil of the digital economy in 21st century, which if refined properly unlocks the potentials of driving optimal and profitable businesses. The effective use of data to optimize business processes has been applied with enormous success in the Internet world and is quickly spreading to other industries. With commercial buildings accounting for over 42% of the energy consumption in the U.S., energy efficiency and reduced maintenance cost are primary objectives in the smart buildings arena. Data science can make a major contribution to improving efficiency.

With the advent of Internet of Things (IoT), cloud solutions, Big Data technologies and smart sensing solutions, novel control solutions are on the rise, along with a rapid increase in the amount of data generated in buildings, as building managers realize value in these data beyond conventional controls. For example, in 2013, Microsoft was collecting half a billion records per day from their buildings at their Redmond headquarters. With data-driven decision making and advanced algorithms, the company forecasts six-10% annual energy savings in these buildings, which are already optimized with smart controls. Similarly, the General Services Administration (GSA), with their GSALink project is collecting 27 million data samples per day from across over 13,000 pieces of equipment.

Navigant Research (formerly Pike Research) issued a report last year estimating that the size of the smart building managed services market alone will grow from $291 million US in 2012 to $1.1 billion US by 2020. IMS Research estimated that the Americas market for integrated and intelligent building systems would be worth more than $24 billion US in 2012. Data science will have a major role in driving these numbers.

As a result of IoT, smart devices, the Cloud, etc., building systems are becoming more complex, made up of more and more disparate devices, all of which contain data and could benefit from the ability to interact with each other. As an example, HVAC devices may utilize Modbus or BACnet protocols while lighting might utilize a proprietary (e.g., Dynet) protocol. The communication among disparate subsystems employing different protocols is a problem that is already being addressed by integration platforms such as Tridium's Niagara. However, as outlined previously, smart buildings still face a huge challenge as data-driven building applications try to mine through and interpret the enormous variety of data generated by a variety of building subsystems. As there are no unified semantics, current building applications are not portable, scalable and almost always require a human in the loop. Due to the lack of a common language and data modeling approach, building data remain in silos. There is therefore a need for normalization of descriptive information about data (metadata) so that the true potential of data can be unlocked. Such a data modeling solution will enable advanced applications to autonomously understand what each data point means what each device is doing and how to use the data.

Metadata modeling is becoming more important in the building automation business as the industry is accelerating towards coordinated control, distributed intelligence, centralized management and data driven services. Metadata and taxonomy are two important, related, but different terms in data analytics. The term metadata stands for descriptive information about a particular datum, in other words metadata are "data about data". For example, consider a data point named "Occupancy", generated by the lighting system in a building. For real-time control and data analytics, we would need more information about this data point such as:

- How to interpret the value of this variable? E.g., 1 – occupied, 0 – unoccupied.
- Where is this occupancy measured? E.g., 3rd Floor, Room #32.
- Which lighting or thermal zone does this point control? E.g., Zone #3.
- All such information can be modeled in a coherent way so that the machines can read and interpret them autonomously.

On the other hand, taxonomy refers to categorizing and relating data. It answers questions such as:

- Where is this occupancy data point originated?
- How is this point related to other data or devices in the building?
- How can one perform mathematical operations on a particular set of data points?

In current practice, metadata are typically represented in the naming convention of a variable and kept internal to a particular system (say a lighting controller). Also, taxonomy[2] is either incomplete or inaccurate in existing building systems. Each system has its own "device trees" or navigation paths that do not accurately represent data taxonomy that can be understood or interpreted effectively outside of that system.

A standardized way of data modeling will free future applications from device or manufacturer specific coding and make them portable (plug-n-play) solutions. Taxonomy in building data should enable the ability to create dynamic navigation trees with their device source, physical meaning and interdependencies among devices described in a machine readable form. Such solutions will reduce manual effort and unleash the true potential of data that will lead to a new generation of device agnostic applications.

2.1. The Bigger Picture - Data Semantics and the Web

We have entered an era where data has become a tangible resource that can be directly utilized to create value. While the readers of this paper are likely interested in how operational data from control and equipment systems can be used to improve facility operations, reduce energy costs, and enhance overall operational efficiency, the need to make data across the Web more self-describing and easier to use is a major area of focus by many in the larger Internet community - a community that dwarfs the size of the buildings market.

Understanding how data semantics are affecting digital society as a whole provides helpful perspective on the need to address this challenge within the intelligent-buildings space. The authors believe that understanding data semantics in a wider sense will help our industry move forward in adopting solutions and offer the following suggested reading:

This article on the "semantic Web" provides good context for the discussion about the importance and challenges of bringing semantic information to bear for the Internet in general: http://www.dataversity.net/keep-on-keepin-on.

[2] The classification of things into ordered categories. Ex: "the taxonomy of these fossils". A scheme of classification: taxonomies "a taxonomy of smells".

These quotes, by Brian Sletten in the article highlight some of the key challenges:

"Adoption of a new way of seeing the world is one of the most difficult things to engender. Beyond simple intransigence, technical impedance, sunk cost thinking, fears and incentive mismatch, the dominant paradigms must be exhausted before people are ready to consider a new approach."

"We don't hear about $20 million investments in Semantic Web technologies though. Is it because the technologies are unnecessary? Overly complex? Poorly understood? There may be elements of truth to these complaints, but at a more fundamental level, Semantic Web technologies don't require consensus to use, and technologies that don't require consensus to use don't require $20 million budgets. If I have the capacity to accept your data as a spreadsheet, a relational database, a REST API, an XML document, etc. and extract content from it on my own, then I can do data integration on the Edge and I don't need your permission to do so. To the extent that there have been Semantic Web successes, they have generally been behind-the-scenes successes. They do not require the kind of political and financial capital that these other huge initiatives do, and that makes all the difference in the world."

Wright State University, which has a program focusing on the semantic Web also, highlights the importance of creating technologies to support the "semantic Web" outside of the interests of buildings industry. The program, called Kno.e.sis (Ohio Center of Excellence in Knowledge-enabled Computing), researches, develops, and applies semantic and knowledge-enabled techniques. Information on the program can be found here and is suggested reading: http://www.knoesis.org.

The takeaway - The building management/energy management industry did not invent the concept of data semantics, and is not the only community to understand the importance of addressing this need. Our industry needs to adopt a methodology that fits the unique needs of our device-oriented applications and can be easily utilized by our industry.

3. APPLICABILITY OF PROJECT-HAYSTACK TO DIFFERENT TYPES OF APPLICATIONS

Project-Haystack technology can be used to address virtually any type of "machine-data" and any application. While much of the work of the community to date has focused on developing tags and taxonomies for equipment systems and devices found in building systems - sensors, actuators, control devices, energy meters, and systems that supply and maintain indoor environmental conditions (HVAC systems), the Project-Haystack methodology itself is application agnostic. Recent additions have included solar systems and water meters and the community is actively looking to broaden the reach of the Project-Haystack effort.

The key takeaway is that the Project-Haystack methodology is application agnostic. Addressing new applications and equipment types is simply a matter of involving domain experts from those fields. Below is a use case example from a member of the Haystack community that combines lighting (which is their area of specialization) and HVAC.

3.1. A Use Case for Meta Modeling in Connected Lighting Systems

Philips is demonstrating its flagship *Connected Lighting* solution at The Edge office building (Deloitte) in Amsterdam. The newly constructed building has 15 stories with more than 6,000 luminaires spread across half a million sq. ft. of office space. The Edge building is a true example of Internet of Things where hundreds of lighting control zones and 30+ thermal control zones (HVAC) in the building will be exchanging information about their respective subsystems. For example, lighting gateways (Envision Gateway) will exchange zone occupancy information with HVAC systems (Figure 2, A Use case for metadata modeling) to enable advanced applications. Philips Lighting is employing Niagara JACE controllers as integration devices to allow lighting systems talk to HVAC devices.

Demand Control Ventilation (DCV) dynamically controls HVAC systems based on real-time occupancy measured by lighting system. On the other hand, Demand Response application dynamically controls different building systems in order to save energy consumption and reduce peak load in buildings. These applications require seamless information exchange between HVAC, lighting, energy and other building systems in realtime. The question that arises here is how to model the descriptive information in a standardized machine readable format so that applications are scalable, portable, and device agnostic?

Imagine datapoints named *mtg_occ* (from lighting) and *zone12_mode* (from HVAC) whose current values are 0 and 2 respectively. How could the DCV algorithms automatically know that the two data points represent occupancy state and HVAC fan speed, respectively, of the same thermal zone (a meeting room)? Currently, integrators use naming conventions to embed some information, but such semantic modeling is a poor approach, and has several limitations. For example, an integrator or manufacturer could name a thermal zone in several possible ways such as zoneA, spaceA, thermA, area1, etc. On the other hand, how do we to interpret the values of these variables? Does the value 0 for occupancy variable mean unoccupied or unknown state? Does the value 2 for HVAC mode mean cooling or heating or standby? Several manufacturers have their own way of representing such states. Moreover, existing semantic methods don't define relationships between data items, i.e., associations.

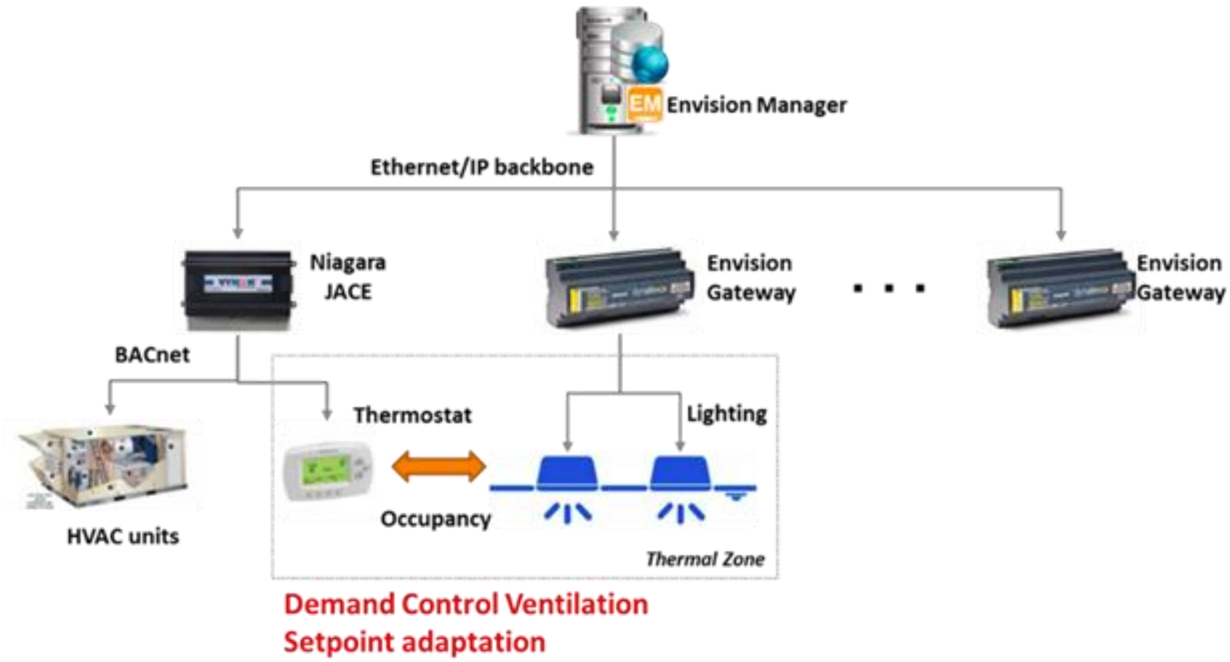

Figure 2 A Use case for metadata modeling

Metadata modeling of building systems introduces a standardized machine readable way to represent descriptive information about data and its relationship. Proper modeling of data points and devices should free applications from the worries described above. Such device agnostic solutions will pave way for quicker deployments, avoid confusion during maintenance and scalable applications.

4. ADOPTION OF PROJECT-HAYSTACK TECHNOLOGY

As an open source initiative where all intellectual property is available for download, the organization does not have a way of tracking individual users or deployments so it's not possible to provide an "official" number of devices or systems utilizing Project-Haystack, but community members report that it is being successfully used in thousands of sites and millions of square feet of building space today. The following information highlights additional market recognition Project-Haystack has received:

2013 Digie Award. Project-Haystack was the recipient of the 2013 Digie Award for Best Intelligent Building Technology Innovation at the Realcomm/IBCon conference June 12-13, 2013. Realcomm has been presenting the "Digie" Awards (short for Commercial Real Estate Digital Innovation Awards) since its first event in 1999. The award recognizes those companies, real estate projects, technologies and people that have gone above and beyond to positively impact our industry through the use of technology, automation and innovation. The award recognized Project-Haystack for:

- Bringing together industry leaders who formed an open source community in order to solve the problem of standardizing the data model for building system data over IP networks
- Dedicated commitment to truly open, interoperable building systems
- Creating a forum for next generation system integrators to define and share best practices

Link: http://www.realcomm.com/advisory/advisory.asp?AdvisoryID=589.

Implementation in Commercial Hardware and Software Products. Project-Haystack technology is being implemented as the core data model in a number of software and hardware products used in building and energy management. Companies that have announced Haystack support built into their products include:

- J2 Innovations
- Intellastar
- DG Logik
- SkyFoundry
- Connexx Energy
- Lynxspring
- Wattstopper
- KMC Controls
- Contemporary Controls

Other public announcements and commentary related to Project-Haystack include:

- Tridium has publicly announced (and demonstrated) their support for Haystack tagging in their major new release "Niagara 4"
- The OBIX committee (a standard under OASIS) has publicly announced that they will support Project-Haystack tagging in version 2.0 of the OBIX standard
- The BACnet committee has publicly confirmed that it is adding the ability to include extended semantic information in its extended data model and the RESTful Web services, and that it is exploring potential use of Haystack tags for an ASHRAE standard dictionary of tags
- Recently, University of California, Berkeley and IBM Research studied existing metadata schemata in commercial buildings. The team has mapped data from three buildings managed by different BMS vendors and consists of more than 6,000 sensors. The evaluation criteria for this study were (a) Completeness of the schemata (published taxonomy); (b) ability to capture relationships needed for applications; and (c) flexibility in modeling. The study group has found that Haystack is the most suitable schemata when compared to BIM (IFC) and Semantic Sensor Web. Source for building systems applications, while also noting that much work needs to be done in this area: [1] Arka Bhattacharya, Joern Ploennigs, and David Culler. "Short Paper: Analyzing Metadata Schemas for Buildings: The Good, the Bad, and the Ugly." Proceedings of the 2nd ACM International Conference on Embedded Systems for Energy-Efficient Built Environments. ACM, 2015
- CABA joined Project-Haystack as an Associate Member and has endorsed the efforts of Project-Haystack
- The KNX Association has joined Project-Haystack as an Associate Member

5. CONSIDERING ALTERNATIVE DATA MODELING SOLUTIONS AND AVAILABLE TAXONOMIES

It is natural that any new standards effort be looked at in comparison to existing technologies that may address the need. As we look to identify alternative approaches for defining data semantics for device data and compare those with the Project-Haystack methodology, it is helpful to start by identifying key characteristics that are important to evaluating different data modeling solutions.

Key Characteristics for Evaluation of Data Modeling Technologies:

- Open Source - freely available to the adopters at no cost
- Methodology that can be used and shared across multiple communication media and data formats
- Human and machine readable
- Extensible methodology that can be adapted to new needs and learning rapidly without long standards approval processes
- Supported by multiple companies and organizations in the marketplace
- Usable by the practitioners involved in deploying controls, automation and sensor systems as well as owners, operators and manufacturers

5.1 Potential Alternative Solutions

5.1.1 BACnet Extended Data Model

The ASHRAE BACnet committee has announced that the soon to be published extended data model (BACnet XD) will include extended semantic metadata, with semantic tags for data and data relations. Siemens Industry, Inc. presented an overview of this approach at the 2015 Haystack Connect conference, as a keynote speaker.

The extended data model will add a rich set of standardized metadata, including semantic tags, not only to its deployed foundation of the BACnet protocol standard, but also to arbitrary structured data, accessible via the soon to be published BACnet RESTful Web services annex will enable full access to this extended data model, with a rich set of query capabilities also on tags.

The BACnet committee approached Project-Haystack for permission to use the Haystack dictionary, with attribution, as one of many semantic tags collections the BACnet committee is considering for creation and future maintenance of an ASHRAE standard dictionary of semantic tags. However, in defining the semantic metadata model in BACnet XD, care is taken to allow for use of tag dictionaries defined outside ASHRAE, e.g., proprietary tags or the Haystack tags as is, and to enable simple and generic translation from the BACnet tagging model to other models, such as Haystack.

- BACnet Extended Data Model: Addenda 135-2012am and 135-2012ba
- BACnet RESTful Web Services: Addendum 135-2012am

Both available at www.bacnet.org.

5.1.2 OBIX 2.0

As mentioned, the OBIX committee (a standard under OASIS) has announced that they will support Project-Haystack tagging in version 2.0 of the OBIX standard, and also expect the standard to support alternative dictionaries that may become available. The authors direct the reader to the OBIX organization for up to date and official information on their metadata efforts. More information can be found at: http://www.obix.org.

5.1.3 OPC UA

A recent article focusing on the ability of OPC UA as it relates to semantic modeling shows the need for data modeling is also a major topic in the industrial automation market and how OPC UA supports data modeling: http://opcfoundation.org/opc-connect/2015/12/why-semantics-matter.

The quote below, from the article, refers to the ability of OPC UA to carry semantic data, but highlights the importance and of developing actual data models (an agreed upon vocabulary) and the work remaining in that area:

"… of the biggest challenges to build intelligent machines (or "agents" in Semantic Web terminology) is not of technical nature. Because "technical concerns" (such as communication protocols, services, infrastructure components, …) are already taken care of by the OPC UA standard and available tools, what remains to be done is the modeling part. In other words, OPC UA offers a language to express and to exchange information, but the "vocabulary" of this language is at least as important if we want intelligent machines to do something useful with this information."

5.1.4 OWL, RFC, RDFS, SPARQL

It's fair to say that the most research and effort in the area of semantic data is being done by the W3C and Semantic Web community. They have produced standards (RDF, RDFS, OWL, SPARQL) that address the challenges of representing, storing and querying information models. Today, Semantic Web technology is used by organizations in a wide variety of industries, essentially to make the best possible use of the vast data stores that they contain.

The above mentioned technology standards could be applied to the buildings and IoT industries, however, they have historically come out of the efforts to bring semantic modeling to the World Wide Web of document-oriented data (versus machine-oriented data). In order to enable widespread use to address the needs of the buildings and IoT markets, a vocabulary would still need to be developed. This is a key part of the value that the Project-Haystack Community has created to date - the vocabulary of community defined and accepted tagging conventions.

5.1.5 Related Links

- http://www.w3.org/standards/semanticweb
- http://www.w3.org/RDF
- http://www.w3.org/2001/sw/wiki/OWL
- http://www.w3.org/2001/sw/wiki/SPARQL

6. INCREASING AWARENESS OF PROJECT-HAYSTACK TECHNOLOGY

Project-Haystack enables increased access and usability of device data, enabling improved sustainability practices essential to the future. If we take snapshots over time to show us what amount of system measurement is in use for building systems we would see an adoption trend that is directly tied to an increasing need for awareness, management and control and a society-wide revolution in understanding the value of data in all human endeavors.

If we increase the awareness of how our built environment is operating and interoperating with the occupants and the external environment, the more compelled and enabled we become to effect improvement. Building systems when designed smartly are able to automatically deliver agreed upon outcomes while also giving its occupants and operators some desired level of awareness of the operating ranges and impacts of the system. Project-Haystack methodologies and protocols will ensure that this adoption continues and accelerates, which will lead to improvements across the buildings and IoT industry.

Project-Haystack helps to ensure the completeness of building system measurements and to enable us to make better decisions that improve efficiency and financial performance. There are many steps between the physical measurement and an improved outcome. Project-Haystack is an effort to provide that agreed upon set of conditions that will enable us to drive continuous improvement. Driven by a greater awareness of our built environment, we can show how increasing and improving measurements within a building and its systems leads to improved outcomes.

Stories sell and facts tell. More than a snapshot, a story or narrative inspires change for the better. This is the heart of persuasion in marketing. Project-Haystack enables data to be more effectively used to provide analytics that are automated, continuous and visual, allowing the story of positive outcomes to be clearly conveyed to building occupants, operators, owners and financial managers.

Today, there are many examples of the results achieved using Project-Haystack. The Project-Haystack organization is dedicated to driving awareness of these successes through participation in industry events, speaking engagements, social media, and market outreach and limited advertising. As an open source, non-profit organization, the financial resources of Project-Haystack are limited. This makes it incumbent on the community to share their success stories and contribute to the ongoing educational process.

7. LEARNING MORE ABOUT PROJECT-HAYSTACK THROUGH FREQUENTLY ASKED QUESTIONS

Q 1: Many small, low power devices are resource constrained. Can Project-Haystack be used in systems that employ those types of devices?

A: It's essential to emphasize that Project-Haystack is first and foremost about defining a methodology to express metadata for device data - it does not require that it be used in the smallest end devices for it to provide benefit. The "Haystack methodology" can be used across media literally from pencil and paper to spreadsheets to Web services and binary protocols. Project-Haystack enables the semantic modeling to take place at whatever level of the architecture makes the most sense and can be supported by the capability of the device. This graphic is illustrative of the point:

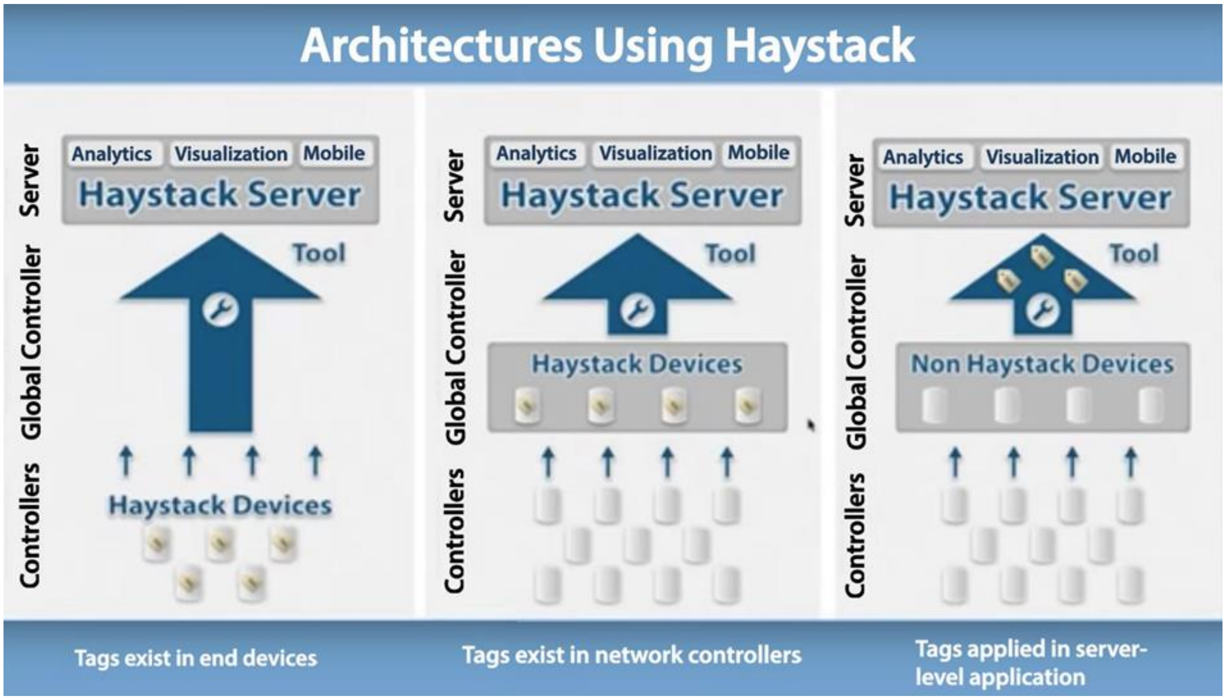

Figure 3 Project-Haystack semantic modeling at various architecture levels

While the ideal "end state" is one in which all end devices include metadata to make their data "self-describing", it is not necessary to achieve that end state for Project-Haystack to provide substantial benefit (First graphic panel in Figure 3).

As shown in the second (middle) graphic panel above, if the end device can't (or doesn't) support tagging internal to itself, tagging can be handled at the network controller/global controller level, which can hold necessary metadata information for all downstream connected devices. The global controller then provides metadata to other devices and applications. If the network controller can't support tagging, it can be provided at the host application layer (graphic panel 3).

The point is to add metadata at the lowest level practical so that layers above can then understand, interpret, share and use device data with far less (ideally no) human intervention. This dramatically reduces engineering costs for BAS, energy management, analytics, reporting visualization and similar applications.

Further to the concern about the capability of end devices, in much the same way the early generation BAS systems implemented BACnet over time, starting in network level controllers before adding it to smaller devices on MSTP RS485 networks, we believe that Project-Haystack will follow a similar path.

Further to the concern that end devices may require more memory and power to utilize Project-Haystack, our experience is that this is not a barrier with any modern devices, or ones in the manufacturer's development pipelines, even down to sensor level devices. The Project-Haystack methodology is extremely lightweight, can be implemented in C++, java, node.js, dart and others technologies, and makes minimal impact on memory or CPU horsepower. The lightweight design and inherent flexibility of the Project-Haystack methodology was intentional, as we are very aware that control devices and sensors have less capability than PCs, smartphones and similar devices.

Q 2: Will the adoption of Project-Haystack technology in end devices make network management and field maintenance become more complex?

A: Project-Haystack does not make any demands on how network management is performed or have any effect on field maintenance of devices. It's simply some additional information held by a device that describes what its points mean to the outside world. It is reasonable to argue that by providing a standardized and consistent method for devices to hold more complete semantic information about their points and usage, Project-Haystack will make it easier to setup and manage devices over time.

Q 3 : Do field devices need to be field programmable to take advantage of Project-Haystack ?

A: Project-Haystack tags do not require field programmability. Tags can be added in simple configuration screens, and/or provided pre-defined by the manufacturer for application specific devices. For example a VAV controller, could provide all of the tagging defined in the Haystack VAV model ready to go. So, while programmable devices can utilize Project-Haystack and allow the programmer to add tags to define that meaning of the various inputs, outputs and variable used in customized control sequences, Project-Haystack by itself does not require a device to be field programmable.

Q 4: Doesn't the addition of tagging mean that communication networks will be required to transmit more data?

A: There are really two aspects to address regarding this question. First while Project-Haystack adds more information, the amount of data that is transmitted on the network is affected by the overall protocol design. The additional amount of data involved in supporting Haystack is overcome with more efficient protocol designs.

Q 5: Will more complex communication protocols be needed to communicate the additional tagging data defined by Project-Haystack?

A: The protocol implementation of Haystack is very simple, clean and efficient and takes advantage of the very latest design patterns and techniques. It's also important to re-iterate you can utilize the Project-Haystack tagging methodology without implementing the Project-Haystack protocol. It's a method to "mark up" control and equipment data in a uniform way so that it can be understood by other people and applications. While the Project-Haystack protocol (and reference implementations that are available in Java, C++, Dart, node.js) were developed by the community to help other community members quickly implement Haystack-based communications in their devices and applications, the Haystack protocol is separate from the tagging methodology. The Project-Haystack protocol is very simple and lightweight, which has led to its adoption, but you can implement and communicate Project-Haystack tag information in other protocols. For example, the BACnet committee is designing semantic tagging extensions to BACnet in the extended data model and RESTful Web services. Those extensions could carry Haystack tags.

Q 6: The Project-Haystack tagging libraries don't currently include tags for my specialized application. Can I still make use of Project-Haystack?

A: The Project-Haystack methodology can be used beyond just the consensus approved tags available at any point in time. Projects and products can add custom tags to represent information important to their needs without requiring submission or approval. And those custom tags can still be discoverable and interpretable by other applications with minimal effort to add a reference index or look up table. Applications that are "Haystack tag-aware" can be easily extended to interpret new community approved tags and custom tags as they are added.

It is also important to note that tagging libraries will always need to be extended. Buildings systems are highly non-standard and new applications, smart devices and other products are constantly being developed. The Project-Haystack methodology excels at allowing for rapid extension of tagging libraries as new application needs are identified. The reach of Project-Haystack for IoT and smart devices is limited only by the involvement of the community.

8. PROJECT-HAYSTACK CORPORATION - STATUS UPDATE

Project-Haystack.org was founded in March of 2011 by a group of people that saw a major unfilled need across the markets involved with smart devices and equipment systems. Today's automation systems, equipment, metering systems, smart devices and IoT applications produce tremendous amounts of data. This data can be very hard to organize and use across different applications because it is stored in many different formats; has inconsistent naming conventions and very limited data descriptors. In essence, it lacks information to describe the meaning of the data. And without meaning, a time-consuming manual effort is required before value creation can begin.

Taking cues from other data intensive applications such as Facebook, Twitter, Google and others who utilize tags and semantic data models, the work developed by the Project-Haystack community addresses this challenge by defining an easy to use methodology to describe the meaning of data from smart devices of all types. This enables software applications to automatically consume, analyze and present data from devices and equipment systems. The open source Project-Haystack effort has streamlined the interchange of data and the techniques for managing, presenting and analyzing the vast amount of data generated by today's smart devices and building and energy systems.

The community develops standardized semantic data models and Web services, with the goal of making it easier to unlock value from the vast quantity of data being generated by the smart devices that permeate our homes, buildings, factories, and cities. Applications include automation, control, energy, HVAC, lighting, and other environmental systems.

The Project-Haystack effort is an open-source community-based effort. Today there are over 1,000 registered members of the community. To better support the growing effort, the community established a formal 501C Corporation with a board of directors from industry in June of 2014. The press release announcing the 501C Corporation can be found here: http://www.ireachcontent.com/news-releases/project-haystack-announces-formation-of-non-profit-corporation-263428181.html.

In addition to the community driven standard work, one of the major activities the community undertakes is the production of the Haystack Connect conference. The inaugural event was launched in May of 2013 and was a tremendous success. The community overwhelmingly voted to hold the event again in 2015 (The event is on a two-year cycle).

At that first event, keynotes addresses included speakers from: Autodesk, the U.S. Department of Energy, and the Governing Institute. Attendance was just under 200 people with 26 exhibitors and sponsors, plus a range of media sponsors. The 2013 Web site is still available here: http://2013.haystackconnect.org.

The 2015 Haystack Connect event was similarly a resounding success. Attendance increased over 45% to 254 attendees. Keynote speakers included:

- **David Mantell**, Global Solution Architect – M2M Americas, Vodafone Global Enterprise.
- **Milan Milenkovic**, Principal Engineer at Intel's Internet of Things Business Group.
- **Jack McGowan**, an internationally recognized expert in the fields of Energy Services, Efficiency, Building Automation, Analytics and Smart Grid.
- **Peter Kelly-Detwiler**, Former SVP of Constellation.

Full details on the 2015 event along with most presentations can be found here: http://haystackconnect.org.

This video provides a two minute overview of the vision of the Haystack Connect event: http://haystackconnect.org/video/haystack.webm.

8.1 Associate Membership Class

One of the major announcements to come out of the 2015 Haystack Connect event was the establishment of a new Associate membership level in the Haystack organization. Response has been very positive and a number of organizations add their support by joining as Associate members:

- Altura Associates http://www.alturaassociates.com
- BUENO Systems http://www.buenosystems.com.au
- Building Systems Solutions http://buildingsystemsolutions.net/bss
- CABA http://www.caba.org
- Controlco http://www.controlco.com
- Grosvenor Engineering Group (Australia) http://gegroup.com.au
- Intelligent Buildings http://www.intelligentbuildings.com
- KNX Association http://knx.org/knx-en/index.php
- KMC Controls http://www.kmccontrols.com
- BASSG http://bassg.com

These new companies joined the Haystack founding member companies, which make up the Board of Directors of the 501C Corporation:

- Airmaster (Australia) http://www.airmaster.com.au
- Enerliance/Yardi http://enerliance.com
- Lynxspring http://www.lynxspring.com
- J2 Innovations http://www.j2inn.com
- SkyFoundry http://skyfoundry.com
- Wattstopper http://www.wattstopper.com
- Siemens Industry, Inc. http://www.usa.siemens.com/buildingtechnologies

Complete information on Project-Haystack can be found at www.project-haystack.org. You can contact members of the Board of Directors at: projecthaystackinfo@gmail.com.

Continental Automated Buildings Association

888.798.CABA (2222)
613.686.1814

Connect to what's next™

www.caba.org