Transforming Markets with Data Transfer Standards

Richard Andrulis and Greg Thomas, Performance Systems Development,
Robin LeBaron, National Home Performance Council

ABSTRACT

Home Performance XML is simple in concept – a national standard to easily communicate home retrofit data. The Building Performance Institute (BPI) is spearheading an effort to define this standard. Much as gbXML enhanced efficiency within the Building Information Modeling field, HPXML has the potential to transform the home performance industry by increasing productivity and reducing labor and overhead costs. HPXML also has the potential to open the market for software within the industry. In the past, many whole house programs have adopted and required their contractors to use vertically-integrated software systems that neither fully address contractors' business needs nor communicate well with other parties. By facilitating communication between systems, HPXML will enable innovative new software products to enter into the market.

The new national standard is designed to support the needs of home performance programs, quality assurance agents, and financial institutions. HPXML can also aggregate data across programs and remove redundancies in incentive compliance to support these additional consumers.

This paper discusses the benefits of information standardization including the possibilities for transforming both the marketplace for auditing tools and the program quality assurance process. We include lessons learned from pilot integration projects including balancing the need for both standardization and flexibility in data definitions and achieving an optimal level of detail. During the alpha testing phase, challenges and issues emerged related to how HPXML describes improvements to a building. This paper describes these challenges and proposes some preliminary solutions that will be taken up as the HPXML development process continues.

Introduction

Goals of data standardization

Within the past decade, XML has become ubiquitous in the information technology (IT) industry because it allows for easy transmission of data in a structured but flexible manner. XML is a standard for describing data so that it can be transferred among and understood by different software systems. The backbone of inter-system communication is now the web service with its self-defining language based on XML.

XML standards are crucial for facilitating data transfer among multiple software systems. If only two systems are communicating, they can develop their own unique XML protocol for their specific data transfer purposes. If multiple systems are communicating the same or similar data among one another, however, it is much more efficient to develop a single XML standard than to have each pair of systems develop their own specific data transfer protocol.

As a result, many industries have defined XML standards and have and seen increases in efficiency and new creativity in the marketplace as a result. The example closest to the home performance industry is the Green Building XML schema, or gbXML, which was created to transfer building information stored in CAD building models.

The home performance industry has reached the point that an XML standard is urgently needed. The industry is now at the point of maturation where the need to verify the energy savings achieved through retrofits has become both possible and necessary. Collection of data will also facilitate the research that will lead to new methods for achieving efficiency. HPXML is designed to provide this crucial industry standard.

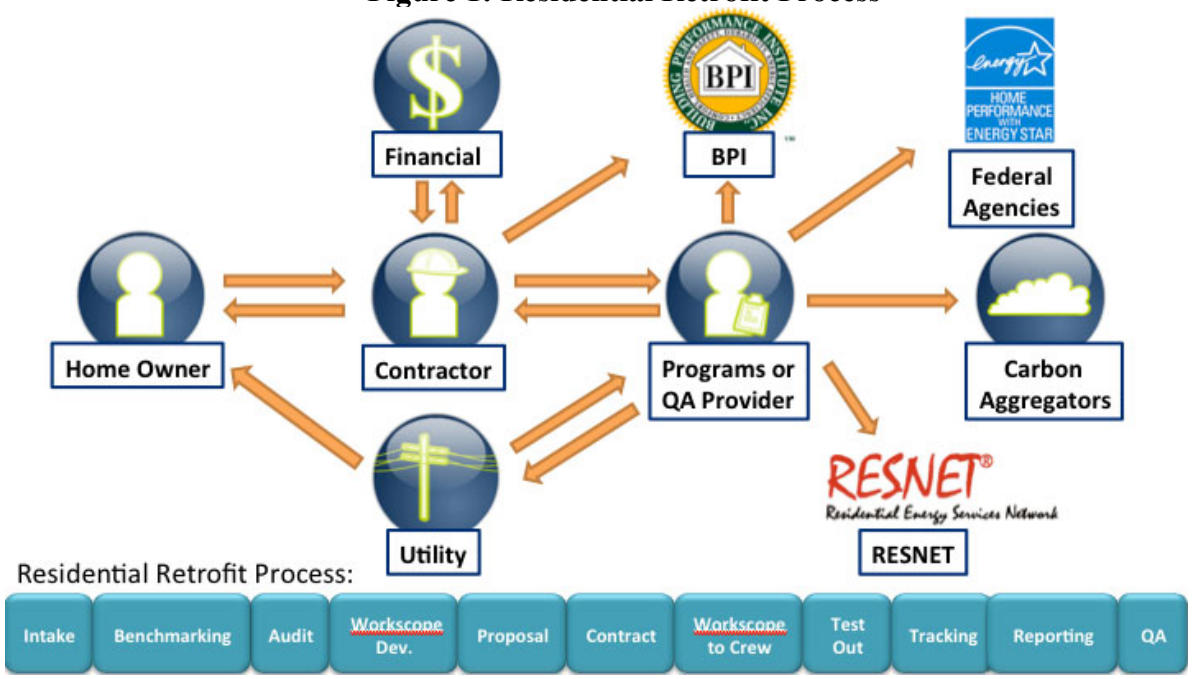
Problems in Home Performance industry

HPXML addresses two data-related challenges in the home performance industry: efficiency and interoperability. Data collection and transfer imposes significant costs in time and IT expertise on both contractors and program administrators (NREL 2011a, NREL 2011b). Reporting protocols frequently require contractors to conduct some "manual" data entry (i.e. reports are not automatically generated by the software). In some programmatic contexts, contractors must submit multiple reports to different program administrators, financial institutions, and other entities. Contractors frequently invoke the time and cost required to carry out this reporting as a drag on profits and a disincentive to participate in home performance programs. Program administrators also incur significant costs associated with the time necessary to collect and check the data submitted by contractors and to reformat it for reporting to program sponsors, such as state energy offices, utilities, and federal agencies. The additional reporting requirements associated with the Better Buildings program have exacerbated these problems.

Home performance programs have traditionally chosen one of several vertically integrated software platforms to provide auditing, data collection, and program administrative purposes. Auditors and contractors are typically required to use this program-approved software to review the results of the audit, model savings, and submit the data necessary for approval of job scopes and financial incentives. This integrated approach reduces the program's data collection and transfer costs, but in doing so effectively displaces them onto the contractors. Contractors need a range of business-related software systems for accounting, sales and other business needs. The software selected by home performance programs typically does not interface easily with these other business-related software packages, requiring contractors to enter data multiple times. Contractors operating in areas served by more than one program may be required to use two or more program-mandated software systems, in addition to their own business packages.

Further, if the program then needs to pass the data on to other parties, custom effort is required to support each additional party. As Figure 1 demonstrates, many parties have an interest in collecting and processing data in the retrofit process. Although it is theoretically possible for each party to develop a customized data exchange protocol, in practice this approach involves a tremendous waste of time and energy.

Figure 1: Residential Retrofit Process



Goals of HPXML

The overall goal of HPXML is supporting the communication between any and all parties that need to transfer data related to whole-house energy efficiency upgrades, as depicted in Figure 1. This is a complex process, as each party typically requires slightly different data for its own specific purposes.

Program/Quality Assurance Provider

Programs are interested in several sets of information including:

- A full assessment of the energy-related components of the existing building;
- A health and safety assessment of the existing building;
- A set of recommended and installed energy conservation and other measures;
- A health and safety assessment of the improved building (the "test-out" data);
- Quality assurance

For market-based programs, there is an additional goal to empower contractors to invest in the auditing tools of their choice, yet still be able to easily report the results of the audit to the program. Often as successful programs mature, contractors become more sophisticated in terms of marketing and sales to ensure their own success as profit making companies. For programs focused on market transformation to support this evolution in contractors' marketing savvy and internal business process sophistication, the programs need to provide a stable reporting mechanism that allows contractors to invest in IT solutions that help them meet their specific

business needs. The use of HPXML as a reporting mechanism will provide this stability, thus enabling contractors to become more successful, and therefore benefiting the home performance industry as a whole.

Financial Institutions

The main goal of the financial institution is to review installed measures to ensure that they qualify for available financial incentives. In some programs, the financial institution may be a single entity, but other programs leverage a network of local banks or credit unions, and thus communication of information in a standard format is critical.

Utility

Utilities are primarily interested in receiving energy savings information, while programs are interested in obtaining energy usage information from utilities. The need for data transfer between programs and utilities will increase as the number of regional programs that operate in multiple utility jurisdictions increases.

For utility-sponsored programs, this is less important, but as more and more regional programs become established, better integration with utilities is essential. Providing a standard format for the transfer of energy savings and usage information will make this integration easier, and allow utilities to invest in the IT infrastructure necessary.

BPI WG-5 has recently engaged in discussions with the team responsible for developing the national "green button" XML standard, designed to transfer utility data related to energy consumption. It is anticipated that HPXML will be designed to align with the green button XML standard.

National Carbon Aggregators

Similar to utilities, carbon aggregators are interested in energy savings information that they can translate into savings of carbon dioxide and other greenhouse gas equivalents. In this case, the use of a national standard is essential, so that the aggregator will not need to provide their own unique standard and thus require multiple data formats and interfaces - a non-starter on a national scale.

BPI/RESNET

Similar to aggregators and utilities, BPI and RESNET are interested in retrofit data from contractors across the country. In this case, they are also interested in the health and safety information in addition to the energy savings. Note there are times when BPI provides quality assurance services for "sponsorless" programs. In these cases, BPI would also be interested in the same information as the Program/QA Providers as mentioned above.

Federal Agencies

The sponsors of the Home Performance with ENERGY STAR® programs, the U.S. Environmental Protection Agency (EPA) and the U.S. Department of Energy (DOE), require

these programs to report on several metrics, such as jobs completed. The Better Buildings program has required even more extensive reporting, although this requirement may not outlast the program funding.

History of HPXML

In 2009 EPA launched a "sponsorless" Home Performance with ENERGY STAR program in northern Virginia (NoVA). The NoVA program was designed by the Environmental Protection Agency to test the extent to which contractors in a promising market for whole-house energy efficiency upgrades (northern Virginia) would be willing to sell and report home performance jobs in a context that provided a QA framework, and allowed them to use the ENERGY STAR logo for marketing purposes, without offering rebates or other incentives.
Data transfer rapidly emerged as a challenge in the NoVA effort. In the absence of a program sponsor, the NoVA program was designed to allow contractors to choose their own auditing software, rather than imposing a single vertically-integrated system as described above. Contractors were required to transfer basic data about their jobs directly to the EPA. To receive data from multiple data collection and modeling systems, EPA recognized that it needed a way to standardize the data transfer process. Performance Systems Development (PSD), which was supporting the NoVA effort at the time, proposed the idea of an Adobe PDF form backed by an XML schema. The contractor would enter data into the form and uploaded it to a website, allowing extraction of the data necessary for programmatic reporting via XML. PSD worked with the Building Performance Institute (BPI) on the data elements needed to ensure the contractor collected a complete set of data. This effort was the genesis of the Home Performance XML standard.

In summer of 2010, the National Home Performance Council (NHPC) convened a national stakeholder group to develop model forms for the proposed federal Home Star legislation. BPI participated in this process and recognized the potential for using the model data set that the stakeholders developed to help standardize data collection throughout the home performance industry. In the fall of 2010 BPI launched a new working group tasked with developing complementary data collection and data transfer standards. It was assumed that the data collection standard would provide a foundation for the data transfer standard (HPXML) by identifying the data elements to be incorporated into HPXML, as well as specifying relationships between the data elements.

BPI Working Group 5, which brings together more than thirty software developers, home performance program staff, program implementers, and researchers, completed a draft data collection standard in the spring of 2011. Using the schemas developed by PSD as a starting point, the working group developed a set of XML schemas during the summer of 2011. In the fall, three teams began alpha testing the software: PSD and Recurve in the context of the Greater Cincinnati Better Buildings program; Clean Energy Works Oregon, Conservation Services Group, and Energy Savvy in the context of the Oregon Clean Energy Works and Home Performance programs; and NREL for its research and analysis-related data transfer purposes. The alpha testing phase was completed in the spring of 2012. Beta testing began in Oregon, in Northern Virginia, where the NoVA program is now sponsored by the Charlottesville VA-based Local Energy Alliance Program (LEAP), and New York. The Working Group's goal is to complete the standard and submit it to BPI for approval as an ANSI standard by the end of 2012.

Scope of HPXML

Business processes

HPXML is designed to describe the data related to each stage of a whole house energy efficiency upgrade.

Customer intake

During this process, a minimal set of information is collected about the building, including the contact information of the customer/homeowner to enable follow-up by the program.

Building Assessment/Audit

This is the fundamental process (along with Test Out) that HPXML is designed to support. The goal is for an auditor to easily be able to communicate the results of an audit, including both the energy-related and health-and-safety-related information for the existing building, proposed improvements, and associated energy savings.

Proposal

This is essentially a streamlined version of the Audit, with only one set of proposed improvements reflecting what the customer has actually agreed to have installed.

Test Out

The Test-out process consists of two parts, an update to the proposal reflecting what was actually installed, and a set of health-and-safety measurements for each test. Note that BPI protocol states that the contractor should perform a set of heath-and-safety tests each day that they are working on combustion appliances in the home, so there could be more than one set of tests for a job.

Data captured

The fundamental approach of HPXML is to capture data collected at the time of Test In/Out. This translates to the XML structure that essentially defines the building information multiple times for certain events (e.g. Existing Building, Proposal 1, Actual Installation). This is an approach similar to that taken by the NREL National Residential Efficiency Measures Database, which defines components and the range of costs associated with going from one component to another. However, recognizing that one of the fundamental use cases for HPXML is the submission of audit and installation data for review by a program for incentive qualification and quality assurance, there is a strong need to not only indicate the current state of the building, but also how it got there.

Structure

During the development of HPXML it was quickly realized that given the number of processes covered and the number of data fields being captured, there needed to be a structure to the schema that would streamline the process for generating and receiving an HPXML document. The approach chosen was to define individual schemas for specific events in the business processes documented above. The first pass yielded the following:

1. Maintain Customer
2. Maintain Contractor
3. Maintain Building
4. Maintain Retrofit
5. Maintain Utility
6. Report Test Results

These each leverage common elements defined in the Base Elements and HPXML Data Types schemas. Some of the common elements include Building Details (see Figure 2), Utility Details (see Figure 3), Customer Details (see Figure 4), and Health and Safety Test Results (see Figure 5).

Figure 2: Building Structure Details

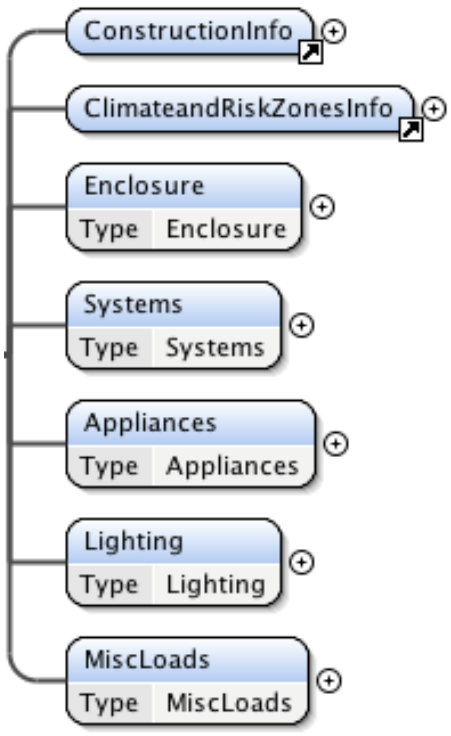


Figure 3: Utility Details

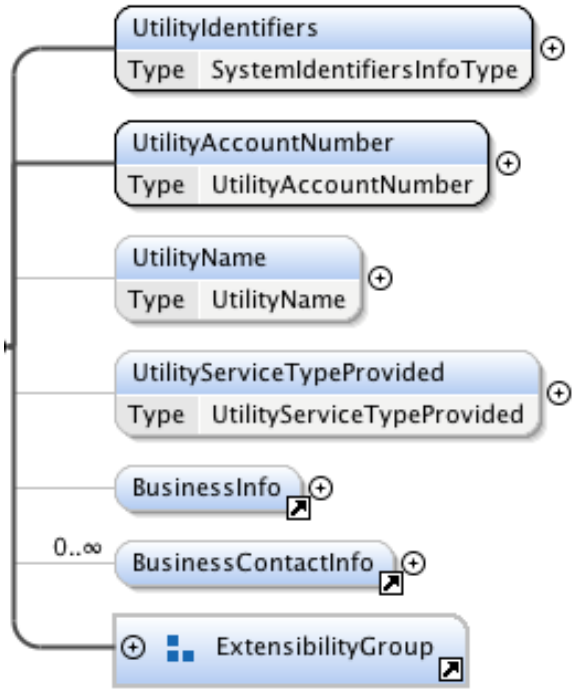


Figure 4: Customer Details

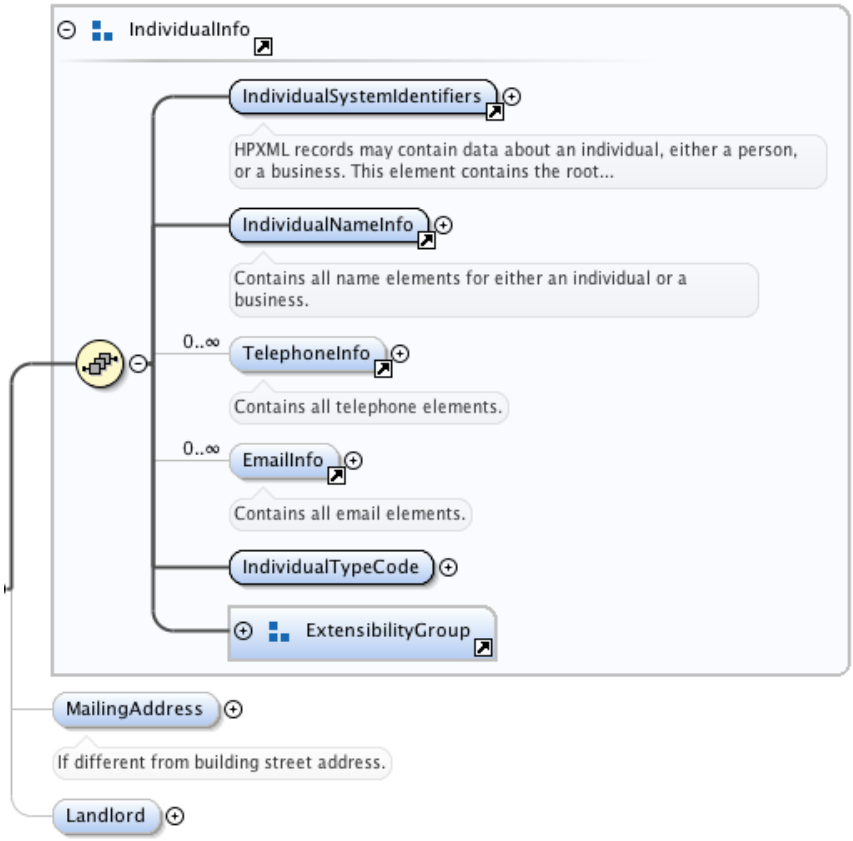
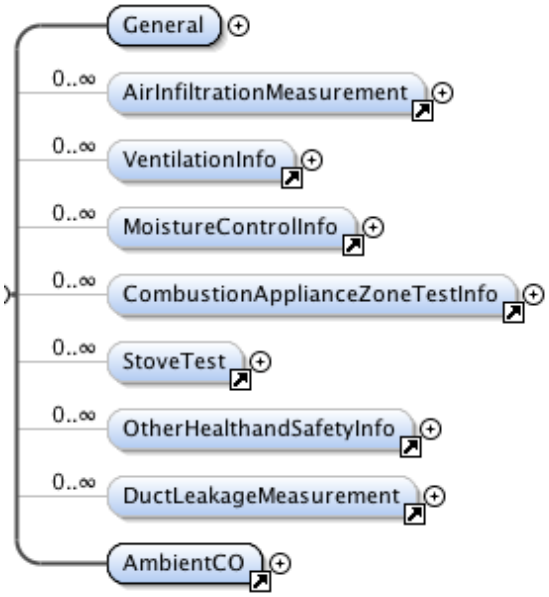


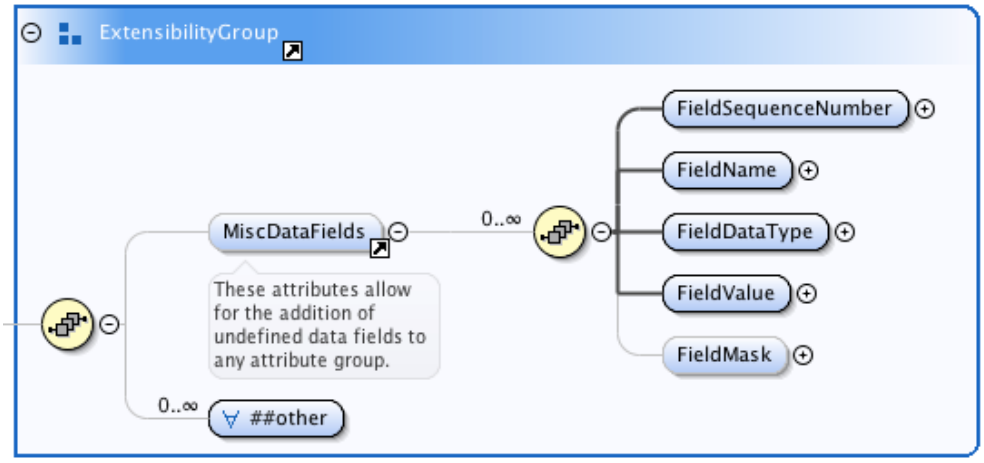
Figure 5: Health & Safety Test Results Details



Several common elements throughout the definition of the standard bespeak the goal of trying to ensure widespread adoption. One such element is the ability to define multiple

Identifiers for one or more sending and/or receiving systems. Another is the ability to flexibly define Events that allow components to be defined multiple times within the same file to capture either multiple proposals, or the pre- and post-retrofit information within the same file. In addition, despite the comprehensive scope of HPXML, it was realized that to ensure adoption by as wide a range of parties as possible, the standard needed to be extensible to support additional, program-specific fields. So, at many of the major nodes, a common extensibility group is available (see Figure 6). This extensibility group provides both a name/value mechanism for supplying additional data as well as the use of a completely arbitrary structure that would allow for the embedding of additional XML, including another standard.

Figure 6: Extensibility Group Structure



Alpha Testing HPXML: Successes and Challenges

Background

As described above, PSD and Recurve agreed to pilot test HPXML in the context of the Cincinnati Better Buildings program. The goal of the pilot project was to allow contractors to use either the PSD auditing tool or the Recurve modeling software to conduct their audit and report the results to PSD's online Green Energy Compass system. Green Energy Compass, in turn, would be used by the home performance program to enable financial and QA review of the proposed improvements and actual installation.

The approach was to leverage Maintain Building to capture both the baseline information as well as the improvements. Recurve was the generating software and Green Energy Compass was the receiving software. To simplify the integration, the approach taken was that the auditor would create the shell of a job in Green Energy Compass and upload a file containing the results of the Recurve audit in HPXML format. Compass would then map the information from the HPXML file onto the job in Compass so that the program administrators would be able to review the audit and job information in exactly the same manner as if the auditor had used the PSD auditing software.

What worked

We were able to transfer details about the existing building information and some simple improvement information. The teams worked through many questions about what data should be filled in where, for example how to handle heating plant efficiencies – basically it was important to explicitly name units. One issue that took some time to get right, but did work in the end was how to structure the parser for the incoming HPXML file. The final architecture was actually a multi-stage parser that first transformed the incoming data into a set of standard components. Those components could then be analyzed to determine how to organize the data (e.g. was there an improvement or not for this component?). While some of the proposed changes to the standard would help alleviate the urgency of this design, the separation of the transformation logic into translation and then organization is sensible and would be a recommended pattern for other software providers to follow.

What didn't work

At the end of the day, the pilot project failed to complete the integration because of several challenges. Both the Recurve software and Green Energy Compass software were designed to describe the building baseline and track improvements, rather than to capture the state of the building at specific points in time. HPXML, as currently designed, allows for the description of the building at specific points in time, and does not define a taxonomy of improvements. The result is that while data coming through HPXML provides value, there are problems.

The "point in time" approach does not translate well to the needs of the Financial and QA reviewers, who need to understand "what changed" so that they can approve incentives for improvements. While it does allow for calculation of the change from the base conditions, a determination of the delta from the base component to the improved component was open to considerable interpretation. For example, if the improved furnace had a higher AFUE than the base component, was this because the furnace was replaced by a newer, more efficient model, or was the existing furnace just tuned up, resulting in an improved AFUE? Things got even more complicated when dealing with situations like walls and identifying which walls were actually improved vs. which had started out with a higher insulation level.

In addition, the concept of streamlining the generation of HPXML by breaking the data into different process steps makes sense in theory; however, the reality is that the most common point for data transfer between two systems is after the audit. At this point, you basically need most of the data contained in Maintain Customer, Maintain Building, Maintain Retrofit, and Report Test Results. If the data transfer mechanism is automated, there is no real issue, but the approach in the pilot required the auditor to download a file from Recurve and upload that file to Green Energy Compass. So, either the generating system has to generate four separate documents, or it has to bundle four separate XML nodes into a single file that can be passed between systems, at which point the benefit of streamlining the generation of the data in HPXML format is lost.

Perhaps the biggest concern was the number of project decisions that needed to be made about how HPXML should be implemented. These ranged from minor issues — such as whether the generating software provides insulation R-value, insulation depth, or both — to more significant issues, such as whether each wall or window should be specified individually, or

grouped into common sets of components. HPXML is designed to be quite flexible; but that flexibility can prove a liability as well as an asset, because it allows each software vendor to make different assumptions about how to use HPXML to describe data. If these differences become significant, the interoperability between software vendors will be greatly reduced, or will fail to materialize altogether.

Lessons learned/Suggested changes

The concept of capturing the state of the building at a point in time makes more sense for an entity like BPI that is fundamentally interested in the state of the building and ensuring that the home is healthy. It makes less sense for programs and financial entities that are as concerned with how the building got to the current state as with describing the current state. One suggestion is to pull the improvement information out of the building structure completely to avoid confusion over what information should be in the Building and what should be in the Retrofit. The proposed solution is to repurpose the Maintain Retrofit message type to act as an index or summary table to the improvements. This index would include information on the "how" (as in "how did the home get from the existing state to the upgraded state") such as a description of the work performed, the cost, who performed the work, etc. The index would also contain a reference to the component details for both the existing building as well as for the upgraded building.

This would allow systems that are primarily interested in improvements to easily pull that information out without having to traverse the entire file to identify components that have improvements. This approach would also address some technical issues related to how to indicate when a component is removed or when multiple existing components are replaced by a single, new component (e.g., replacing several room air conditioners with a central system).

Regarding the concern of how to implement HPXML, much of the risk can be mitigated with proper guidance on what data to provide. We can even see multiple tiers of these that correspond to the level of detail of the audit. A level 0 or walk-through audit would provide the bare minimum of data, a level 1 would provide more details, and so on. These guidelines would be crucial to ensure that an HPXML document generated by one system could be understood by another system without costly rework. Some possible guidelines to start the discussion could include:

1. Use the Maintain Building schema *only* to define building baseline information. While it is possible, and tempting, to cram both base building and improvements into a single file, it causes a great deal of confusion in interpreting the contents of that file.
2. Define Events at the top-most node to make it clear to the receiving system that this set of data belongs together as either the base building or the improved building – effectively the Maintain Building will define a building snapshot.
3. Provide only physically measureable quantities (e.g. depth of insulation in inches rather than the inferred R-value) for Tier 0.

These solutions may be facilitated by BPI Working Group 5 returning to the underlying data collection standard, both to develop a more detailed and explicit language for

improvements, and to provide more guidance with respect to the way that certain building characteristics, such as insulation, are reported. These approaches will be explored through the summer of 2012.

Conclusion

HPXML will benefit the building performance industry in three ways. First, it will allow considerable improvements in efficiency by lowering communication costs. At present, contractors, program administrators, and other parties spend large amounts of time and money transferring relatively limited data sets. HPXML has the potential to greatly reduce these costs by reducing the number of times that data must be entered and by reducing the need to develop transfer protocols between different software systems.

Second, HPXML will allow more and better data to be collected. Data is crucial to the home performance industry, and will become more important over time, as the industry scales up and programs and contractors begin to explore ways to sell savings to utilities, capacity markets and investors. To date, many programs collect relatively little data about the improvements that their programs sponsor; and the data that is collected is not always entirely comparable. By standardizing data collection and facilitating data transfer, HPXML and its companion data collection standard will enable programs to collect much larger and more detailed data sets, which will have multiple research and programmatic benefits.

Third, HPXML will increase interoperability between software programs. This will allow new software vendors to enter previously integrated (and thus effectively closed) markets because programs can be assured that the new software vendors will be capable of providing standardized, high-quality data without requiring the programs to expend excessive time or effort in developing the tools to receive and read their data. By allowing new software vendors to enter the market, HPXML will also encourage innovation and product development. New software programs will be of particular benefit to contractors, who will have more opportunities to mix and match the modeling, marketing, and business software systems that best meet their needs.

While there are improvements in HPXML that are needed to ensure widespread adoption of the standard, we feel that the standard will have a transformative effect on the market. In order to fully achieve these benefits, continued national attention is needed from software vendors, program implementers, and other interested parties. The Working Group (WG-5) created by BPI's Standards Technical Committee has been championing this effort so far; however, there is still a fair amount of work left to do to ensure the widespread adoption of this data standard. This work will be one of the priorities for the working group during the remainder of 2012.

References

National Renewable Energy Laboratory, 2011. *2011 Residential Energy Efficiency Technical Update Meeting Summary Report, Denver*, *CO*, DOE/GO-102011-343 Golden, CO: Building America

National Renewable Energy Laboratory, 2011. *Summary of Needs and Opportunities from the 2011 Residential Energy Efficiency Stakeholders Meeting, Atlanta, GA*, DOE/GO-102011-3283, Golden CO: Building America