
THE FUTURE OF REAL ESTATE INFORMATION

by John A. Kilpatrick

EVOLUTION OF REAL ESTATE STANDARDS

"I don't see information technology as a stand-alone system. I see it as a great facilitator. And maybe more important, it is a reason to keep asking yourself the question – why, why, why."

- Paul O'Neill, former Chairman and CEO of Alcoa,
now U.S. Treasury Secretary¹

For real estate counselors, application of information technology would enable projects to be done faster, cheaper, and with less risk. Consider the typical, simple acquisition: assuming a willing buyer and seller, with no particular barriers to the transaction occurring, the due diligence has traditionally taken three or four times as long as the contracting itself. However, in the fast-paced acquisition environment of late, particularly fueled by the REITs, the period between signing the letter of intent and the closing has shrunk to as little as 10 days.² If due diligence efficiency could be enhanced, even marginally, the cost savings per transaction would be enormous. Given the current estimates of the size of the U.S. real estate market—\$4 trillion at last count—the aggregate cost savings will be huge.³

Consider also the after-the-transaction needs for information—reporting, periodic revaluation, tax compliance, and a host of other data needs, both on the subject property as well as on comparable and competing properties. As real property is increasingly securitized or included in institutional portfolios with quasi-securitization reporting requirements,

ABOUT THE AUTHOR

John A. Kilpatrick is a partner and senior analyst with Mundy Associates LLC, of Seattle, WA. Kilpatrick is a frequent speaker and the author of four books on real estate, most recently *Subdivision Development* published by the Realtors Land Institute.

the need for timely, accurate information becomes a mandate, not an option.

Couple this with the increasingly sophisticated analytical tools. Counselors who fail to use appropriate analytical tools are at the very least planning their own retirements, and are often guilty of failing their fiduciary duties. However, the simple act of transferring data as it may be available into an analytical data set is horrendously time-consuming and fraught with errors and interpretive problems.

Finally, there is the issue of listing data. Internet-based commercial listing services such as Loopnet have, relatively overnight, become the 5,000 pound gorilla of commercial real estate listing information. As an example, in October, 2000, Loopnet had over \$83 billion in for-sale listings and over 1.3 billion square-feet of for-lease listings. Further, Loopnet has emerged as a transaction player as well, featuring over 27,000 purchase and lease "requirements" listings.⁴ Savvy commercial brokers now find that they can do an increasing number of deals just through the Internet. For example, Dean Cruci, a Southern California broker, was recently quoted as saying, "Of the last 10 escrows I've closed, nine out of 10 of them have been on the Internet."⁵

In the early 1990s, the National Council of Real Estate Investment Fiduciaries (NCREIF), the Pension Real Estate Association (PREA), and the National Association of Real Estate Investment Managers (NAREIM), together took the first step by creating the Real Estate Information Standards, or REIS. This project was stimulated by the information shortcomings of the real estate field, relative to other investment segments, such as equities and fixed income. Those investment segments are primarily public in nature, and have had many years of scrutiny and guidance by regulatory bodies. That scrutiny has served to increase the confidence of investors and investment analysts that information is conveyed accurately and consistently. Real estate, on the other hand, has been primarily a private market activity until recently, with regulation and information exchange perceived as being less than complete.⁶

The REIS effort culminated in the publication of the first phase of the standards in 1993, and the selection of Deloitte & Touche LLP to provide ongoing monitoring and improvement.

INFORMATION STANDARDS & THE WEB

It was an interesting juxtaposition of timing that

In the late 1990s, it became obvious to many researchers that there was a gap between REIS and the way information was organized on the Internet. A collaboration was formed, including many of the REIS sponsors, as well as representatives from private sector firms (REITs, accounting firms, major real estate firms, and the evolving set of real estate data providers themselves) to systematically integrate real estate information, under the rubric of REIS, into such standards as would be necessary for full utilization of the new generation of information technology resources. The collaboration came to be known as the Data Consortium (DC).

REIS was published just as the World Wide Web was being created. Even though the Internet had existed for almost two decades prior to the early 1990s, it was mainly used for e-mail among research laboratories and universities as well as transfer of fairly large, specialized research data sets among laboratories. To put matters in perspective, when the new President of the U.S. was inaugurated in 1993, there were about 60 "Web" sites in existence, almost unknown outside of a small fraternity, and powered by a now defunct software package developed by supercomputer researchers called "Mosaic."

In the ensuing eight years, the Internet revolution has been both profound and ubiquitous. Ignoring for a moment the e-commerce applications, and focusing only on the data issues, a researcher in Australia or Austria or nearly anywhere in Asia can share data with colleagues in New York, Old York, or Yorktown as quickly as dialing a telephone. Data transfer is nearly free (certainly cheaper than sending an equivalent sized fax, overnight package, or even first-class letter), nearly 100 percent accurate (what you send is almost always exactly what the other side gets), and potentially, immediately usable. For example, if I am working on a spreadsheet analysis in a hotel room in Dothan, Alabama, and need to share that with a colleague in Seattle, Washington, I can not only send the information, but in a

format that allows the colleague to immediately re-analyze the data without any additional data entry or data manipulation.⁷

Unfortunately, this particular example, and all examples like it, require “two sided” manipulation of the data. In other words, there has to be both a sender AND a receiver. Additionally, the receiver must be using the same data analysis tools as the sender, or at least some tools which are congruent.⁸ Finally, the Web has enabled a huge industry in online databases, enabling an increasing amount of “one-way” traffic in data. Hence, the sources of data (that is, the online databases) must be constructed and designed in such a way to facilitate simple, immediate use of the data in common analytical tools.

EVOLUTION OF THE DATA CONSORTIUM

In the late 1990s, it became obvious to many researchers that there was a gap between REIS and the way information was organized on the Internet. A collaboration was formed, including many of the REIS sponsors, as well as representatives from private sector firms (REITs, accounting firms, major real estate firms, and the evolving set of real estate data providers themselves) to systematically integrate real estate information, under the rubric of REIS, into such standards as would be necessary for full utilization of the new generation of information technology resources. The collaboration came to be known as the Data Consortium (DC).

The DC’s agenda was to find ways to utilize the Internet to address 13 specific problems faced by the real estate industry:

- identifying the commercial real estate inventory nationwide;
- locating and maintaining listings of properties for sale;
- locating and maintaining listings of space for lease;
- performing investment underwriting;
- conducting pre-purchase due diligence;
- aiding transactions/transfer of ownership;
- performance reporting to clients, consultants, data vendors, and trade associations;
- integrating disparate property management systems and information;
- compiling operating or occupancy statistics;
- combining property and demographic/economic data;
- facilitating document storage and retrieval;

- fact-finding for risk management; and
- custom reporting for various clientele.

One of the first steps, interestingly, was to determine exactly what those information technology resources were. In the earliest days of the Web, a common language evolved known as *hypertext mark-up language*, or HTML for short. From the user’s perspective, the most common manifestation of HTML was the coding for Web pages. However, HTML also provided a template within which data was encoded and hence transferred on the Internet.

While HTML is a good tool for developing pretty Web pages, it unfortunately lacks the structure for efficiently and effectively encoding data. For example, both a street address and a sale price can easily be encoded within an HTML document, but both are essentially coded as text strings. Thus, the HTML document itself provides no clues as to which characters represent a price and which represent a street address. For that matter, a zip code and a price look exactly alike within an HTML document.

Clearly, this unsophisticated level of coding would be insufficient as the Web increasingly became the worldwide transfer mechanism of choice for nearly all forms of data. Hence, the second generation of Web coding quickly took form as the *extensible mark-up language*, or XML. This language provided for the necessary tags to encode data by type, not just as text strings. Within XML, a data set could be sufficiently organized to be useful to a variety of users without the need for complex deciphering of HTML documents. The integration of REIS into the technology age would occur under the rubric of XML.

In the end, the Data Consortium (DC) was formed to bring these issues together under one umbrella. Principal coordination support for the Data Consortium is provided by RREEF, while day-to-day work of the Data Consortium is done by a volunteer working group.⁹ The first organizational meeting of the Data Consortium took place in Chicago at NCREIF headquarters on February 10, 2000. Attendees included members of the National Council of Real Estate Investment Fiduciaries (NCREIF) and representatives of the International Council of Shopping Centers (ICSC) and the National Association of Real Estate Investment Trusts (NAREIT). Attendees concurred that the creation of Internet-savvy standards for the interchange of real estate information was an essential task for the group.

The Data Consortium was formally incorporated as a not-for-profit entity on October 12, 2000. More specific information on the Data Consortium and its working group is available at the Web site, <http://www.dataconsortium.org>.

Further helping to define this integration was the predecessor development of the Real Estate Transaction Standards (RETS), promoted by the National Association of REALTORS (NAR). The work of the DC was further complicated with the need to be congruent with parallel efforts of the Worldwide Web Consortium (W3C), the Mortgage Bankers Association (MBA)¹⁰ and the American Institute of Certified Public Accountants (AICPA). Additionally, standards development has proceeded globally, most notably in the U.K. through the Property Information Systems Common Exchange Standard (PISCES), further adding to the coordination burden.¹¹ Finally, since so much of real estate information is also legal information, it quickly became apparent that the work of the DC would have an impact on parallel work being done in the legal community, specifically LegalXML.¹²

The mission of DC was clear—to realize a new class of software oriented to one, vendor-neutral, common information standard. It was widely agreed that without this, the major consequence to the real estate community would be a continuation of the legacy business model based on ‘captured’ customers whose data are stored in disconnected, restrictive computer systems. Without a unified standard for interchanging information, current vendors would have little incentive to improve their products beyond developing Internet interfaces. Collectively, the industry would miss significant opportunities as Application Service Providers (ASPs) create new and exciting applications for which there would be few customers willing to duplicate/straddle their operations. Real estate standards likely would not result by consensus among all the stakeholders, and the industry’s costs to move from a paper-chasing world to an information-chasing world will be far higher than they needed to be.

These efforts are resulting in a set of information interchange standards within XML which has come to be known as the Data Consortium Namespace (DCN). This body of standards will, when published, provide a template that will allow data developed under REIS to be easily shared and utilized among constituent groups.

The mission of DC was clear – to realize a new class of software oriented to one, vendor-neutral, common information standard. It was widely agreed that without this, the major consequence to the real estate community would be a continuation of the legacy business model based on ‘captured’ customers whose data are stored in disconnected, restrictive computer systems.

DATA CONSORTIUM NAMESPACE

The Data Consortium Namespace is a combination of XML element definitions plus one or more dictionaries of industry-specific terms referenced by the elements. Together, these are used to represent data considered within the “domain” of the namespace. The DCN is generally distinguished from other industry namespaces by two groups of information. The DCN must identify and describe real estate properties available for sale, lease, or assignment. Second, the DCN must accommodate measurements of financial and management performance regarding these properties. Neither requirement is today fully supported by namespaces publicly available from the real estate industry, or from allied industries.¹³

To accomplish this, the DCN is developing three major definitional standards which will tie together four interrelated initiatives. The definitional standards are:

Guide to the Data Consortium Namespace (DCN)

DCN 1.5 is a broad set of specifications which defines XML elements and attributes for expressing information related to commercial real estate listed for sale, lease, or assignment. DCN 1.5 includes a set of terms that are referenced by elements in other documents.¹⁴

Document Type Definition (DTD) for the DCN

The DTD is a statement of the information schema for the DCN. The DCN applies the latest technologies to the model recently published by the National Association of REALTORS giving a sophisticated, stable, programming environment.¹⁵

Data Consortium Dictionary (DCD)¹⁶

The DCD will contain elements used to define

references, linking these of Java classes or script objects.

These three basic definitional standards will then form the basis for the development of four interlocking and mutually consistent sets of tools for developers:

1. The DCN Software Development Kit (SDK),
2. The DCN Processes for Information Interchange,
3. DCN Model Applications, and
4. Data Consortium Web Services

DCN Software Development Kit (SDK)

The real estate community would benefit from having a common base of software that can be integrated by vendors into their products. The SDK is envisioned to be a single packaging of software and reference material, and includes samples of all documents within the scope of the DC Namespace. Specific software planned for the SDK includes the following:

DCN Object Classes

The DCN is designed specifically to accommodate a modern software development environment. Most recently, these environments have been based on object-oriented software development methodologies. In keeping with this object-oriented approach, the DCD contains object class names that permit use of software objects from the content of a DCN data stream.

The advantages of this approach to software development are most apparent in the ease of inheriting object behavior across classes arranged in a hierarchy. The 'behavior' of an object class is encoded in a language such as Java, Visual Basic, or Perl, and is reached by sending a message to either a locally-used object or a remotely-used object.

The Simple Object Access Protocol (SOAP) is the messaging protocol of choice for cross-machine communications. Accordingly, message names standardized by the DCN can be recorded in the DCD. Thus, the DCN standardizes both object class names and class message names.

The first objective of the Software Development Kit (SDK) is to standardize object classes that track to the 20 resource and attribute element types defined within the DTD. In addition to classes corresponding to each category defined by the DCD, a DCN Resource class will be defined. The

DC Dictionary and Glossary are also designated as object classes.

To support 'push' and 'pull' techniques, classes for DCN data stream 'readers' and 'writers' are required. The 'readers' use objects implied by the DCN data stream, and the 'writers' create DCN data streams from information in the objects.

The specification for each of the DCN object classes will include class and instance message names, and will be specified in a binding-neutral manner (e.g., using the CORBA Interface Definition Language). Subsequently, at least two code implementations will be created: one supporting server-based processing, the other supporting common client environments. At least one of these implementations will use the Java language.

DCN Validation Services

The DCN is designed to require a custom mechanism for an exhaustive validation of a DCN data stream. While XML DTDs and XML schema are adequate for most syntactic validations, a number of syntactic variations allowed and prohibited by the DCN cannot be validated using those conventional methods. Furthermore, validation must address whether undefined dictionary terms are used by elements in the DCN data stream. Validation must be accessible in real-time, and must create a DCN-conforming stream of error messages.

Because some vendors may need terms not defined in the Data Consortium Dictionary, the DCN allows users to create their own custom dictionary. Custom dictionaries may also be subject to validation at the option of the author.

DCN Processes for Information Interchange

DCN E-mail Protocol

A key service for Data Consortium members is the ability to send documents to another party or to a repository using standardized mechanics. This is the input side of the architecture.

This plan proposes an SMTP-based solution for the large percentage of documents that can be e-mailed from a publisher/sender to a repository/recipient. The solution includes a Java program running on a sender's machine that creates an XML stream that the sender can then attach to an e-mail message. The program can also invoke the e-mail client, e.g. Outlook, located on the sender's machine. SMTP-based filter software located on the

destination's mailbox is also required. Its job is to identify incoming mail having DCN attachments, and to invoke software on the host machine to route the document to the proper location.

DCN Transformation Services

The DCN can also be used effectively as a persistent storage format. There is a need for interfaces to other XML namespaces encountered by DC members in order to take advantage of the promise of having a single data stream that can serve multiple needs. For example, the Extensible Business Reporting Language, (XBRL) is used for reporting certain financial information that could be stored in a native DCN format. The Data Consortium will also identify other XML namespaces with which transformations might be required.

Transforms can also be created to take XML data streams and display them in HTML, PDF, text, or other formats. These transforms or style sheets will be included with several of the model applications described below.

DCN Model Applications

Model applications are a useful way of conveying the power and capability of the DCN. The Data Consortium will create and post for public review a series of templates showing real-world applications of the DCN in enough detail to allow software developers to extend the templates into commercial products and services. The most commonly requested model applications are as follows:

Property For Sale or Lease

In discussions with Data Consortium members, it has become clear that finding a simple and efficient means to post information about property for sale or lease on a wide array of Web-based listing services is a high priority. Property owners and their agents wish to have a system that permits them to prepare the needed information once, and to allow others to gather that information without further involvement by the publisher of the information. Similarly, the listing services wish to have choice of the data they collect, the manner in which it is stored, and the frequency with which it is collected and updated.

Thus, a model application will be built around this need. The system is based on the DCN Software Development Kit, thereby allowing storage of information about properties and spaces available for lease. Additionally, a 'robot' will be designed to query those storage sites periodically and to extract

The real estate industry has, for most of its life, utilized highly fragmented, proprietary data sets which rarely communicate with one another. However, a significant change is underway which will provide the opportunity to bring these disparate sources together with a common language and via the commonly available Internet.

information needed by commercial property listing Web sites.

Sites having information about property for sale or lease will be included in the Data Consortium Registry, described below, to make searches by people or robots more efficient.

Due Diligence Transaction Processing

The process of transacting commercial real estate is complicated and time-consuming, involving dozens of steps and many documents. This activity lends itself to the XML standards being developed by the Data Consortium.

This model application will rely heavily on the DCN-specific e-mail protocol described above in the SDK. In fact, this application and the e-mail protocol are essentially the same initiative.

Property/Fund Reporting Server

The mapping of the Real Estate Information Standards (1999) and associated property and fund level information to the DCN provides the foundation for the development of standards for documents submitted routinely to NCREIF by property owners, managers, and investors. Additionally, these data elements are commonly used to report property and fund performance and supporting material to clients/owners and their consultants or trustees.

The focus of this project is to identify all of the information commonly provided in these reports, and to create an XML-based means of conveying that information in a variety of computer-based and hard-copy formats: HTML, PDF, and others deemed appropriate.

Valuation/Cash Flow Forecasting

Real estate investment decision-making

routinely involves some estimate of value. Since the primary technique for estimating the value of commercial real estate requires computing the present value of net revenues derived from complex lease structures, the problem requires the computational power of a computer. The Data Consortium will create a model application that captures the information typically required to perform valuations and develop the means for extracting that information from legacy systems for transmittal via the Internet to service providers and applications that can process the data.

Data Consortium Web Services

Registry

The Data Consortium will maintain on its Web site a searchable list of databases that adhere to the XML standards developed by the Data Consortium. This registry will make the task of locating real estate information efficient for human users and especially efficient for computer robots that are designed to scan for new or updated information on the Internet.

Other DCN interfaces—such as with RosettaNet and CommerceNet—will be reviewed and mapped, if members wish.

Directory of Technical Support Service Providers

As the Data Consortium identifies people familiar with its XML-based standards who can assist in developing applications, these individuals and firms will be included in a searchable database on the Data Consortium Web site.

Member Pages

The Data Consortium will afford sponsors and consortia of any category of members the opportunity to have individual Web pages on the Data Consortium Web site.

Branding

To help promote the Data Consortium and to allow its members to show their support for the standards, the Data Consortium will develop several icons or graphic images that members can display on their individual Web sites.

SUMMARY

In his book, *Business @ The Speed of Thought*, Bill Gates summarizes both the challenge and opportunity of the information age, "Business Leaders who succeed will take advantage of a new way of doing business, a way based on the increasing velocity of

information."¹⁷ Indeed, that is what this, admittedly very technical article was all about. The real estate industry has, for most of its life, utilized highly fragmented, proprietary data sets which rarely communicate with one another. However, a significant change is underway which will provide the opportunity to bring these disparate sources together with a common language and via the commonly available Internet.

Real estate counselors who become fluent in this new schema, and can take advantage of the rapid shifts in information availability will gain almost insurmountable advantages in the market. For example, it is clear that the 30- to 45-day due diligence period will quickly become a thing of the past, and something closer to 10 days will become the norm.¹⁸ Counselors will need to adopt and adapt to this new standard. Listing data, (both on the buy-side and the sell-side), and closing data will quickly move to the Internet, forcing brokers, appraisers, and others dependent on this information to adapt to the changing information world.^{REI}

NOTES

1. Quoted in Gates, Bill, *Business @ The Speed of Thought* (New York: Warner, 1999), p. 295.
2. DeMay, Tracy L., "Real Estate Due Diligence and the Closing Process", *Real Estate Review*, Fall, 1998, pp.45-52.
3. The Commercial Investment Real Estate Institute, as cited in Grebb, Michael, "Hot Property," *Business 2.0*, December, 1999.
4. <http://www.loopnet.com>
5. Grebb, *op. cit.*
6. *Real Estate Information Standards 2000*, (© 2000), NCREIF, PREA, NAREIM, available on the Web at <http://www.ncreif.org/reis.htm>
7. This is an actual, recent example from the author's files. Other examples, of course, are too numerous to mention.
8. For example, data input into a spreadsheet is usually readable by, say, higher level statistical analysis packages, such as SAS or SPSS. Unfortunately, the same is not always true with proprietary packages, databases software, or other tools. Further, this cross-platform manipulation often takes a degree of programming sophistication.
9. Other sponsors of the Data Consortium include Constellation Real Technologies, Pension Advisors Consortium on Technology, Management Reports International, and the Realm.
10. The MBA has taken the lead in developing XML standards for residential mortgages through the Mortgage Industry Standards Maintenance Organization (MISMO). For more information on this, visit their Web site, <http://www.mismo.org/>
11. PISCES is an XML data exchange standard developed for the U.K. property market. Its management and development is being undertaken by a group comprised of suppliers of property management systems and their customers. More information can be obtained at <http://www.pisc.es.co.uk/>
12. For more information on the interface between the DC and LegalXML, visit the latter's Web site, <http://www.legalxml.org/>

13. Much of what follows has been adapted from the original business plan for the Data Consortium, developed by Michael Young of RREEF and Richard Kozak of Management Reports International.
14. The candidate standard, Version 1.5, was released on January 18, 2001.
15. The candidate standard of the DTD is currently available for comment and voting by members of the Working Group.
16. More rigorous definitions of DCN, DCD, and DTD are available on the Data Consortium Web site.
17. Gates, Bill, *op. cit.*
18. DeMay, Tracy L., *op. cit.*