chapter 1 The Data Turn

Model quality is certainly improving, but we are still not seeing enough valuable embedded data.

—David W. Light

Until recently, the discussion of data wasn't a daily occurrence in most architecture, engineering, and construction companies. Why then is there a need today for an understanding of how data is being leveraged in architecture, engineering, and construction, and by owners and operators? In other words: specifically for the AECO industry, why is this happening now?

Five Factors Leading to the Leveraging of Data and Industry Change

What forces and technologies have come together in the second decade of the twenty-first century that make the gathering and use of data possible for industry practitioners in firms small, medium, and large?

Technology

Technology has played a large part in the rise of data availability and use, including increased computer power, enabling the ability to crunch large quantities of data and provide higher-resolution communications, access to the cloud, and less expensive storage options. Software has a role in all of this as well. We have started to ask how building information can be better leveraged using data mining, and have started to investigate new directions for accelerating the flow of building information throughout a facility's life cycle. In turn, we have started to see where BIM data is being used in decision making in design, construction, and building operations. (See Figures 1.1 and 1.2.)

Many design and construction professionals—and also their clients—are justifiably frustrated that promised results from BIM tools are not being more readily achieved. The reason for this delay is that so-called higher uses of BIM—analyses, including scheduling, cost estimating, energy, sustainability, facilities management, and facility operations—require not only collaboration on integrated teams, but also the collection and strategic application of building data. Another factor is higher-resolution communications. Soon people will be able to share vastly more information than they are currently. I asked Andrew Witt, Director of Research at Gehry Technologies, if this can be attributed to an increase in the need to share or something else. "It's the opportunistic availability of both data and the means to share

Figure 1.1: BIM Benchmark measures real-world performance of computer hardware. Users are presented with a series of statistics concerning how quickly their computer executed a series of tasks in a BIM model, allowing them to make more informed hardware-purchasing decisions. © *CASE*

Figure 1.2: A version of the BIM Benchmark tool prototyped at CASE. © *CASE*

it," says Witt. "It's not necessarily based on some new requirement to share. There's a greater and greater expectation of higher and higher fidelity communication. People will have the means to execute high-resolution communication. People won't necessarily be communicating more frequently. But the resolution of that communication will be much higher."

The higher resolution will enable more data and information—and more exact data and information—to be shared more quickly and more reliably. Part of this is being brought about by cloud computing. Mads Jensen, CEO of Sefaira, admits that he wouldn't have a product if not for the cloud: "With cloud computing, we can now analyze everything in far greater detail, thereby using the analysis of our design data to actually shape the next design decision." (See Figures 1.3 and 1.4.)

Strategy No. 3: Look Outside the Industry

The architecture profession and construction industry have always trailed mainstream technology. CASE's David Fano suggests one way to keep up or even stay ahead: "If you want to see what's coming up for the AEC industry, just look at articles in *TechCrunch*[1] from five years ago. You can see where the world is going. If anything, we're behind."

Fano takes a contrarian view, holding everything that appears new today has actually been with us for some time: "How long has business intelligence been around? It's old news. For the AEC industry, it's a new, innovative, groundbreaking thing—it's really not. That's what I tell people—others have figured this out for us already. The technology's figured out. The software's figured out. Processes are mostly figured out. We just have to readapt them to our industry."[2]

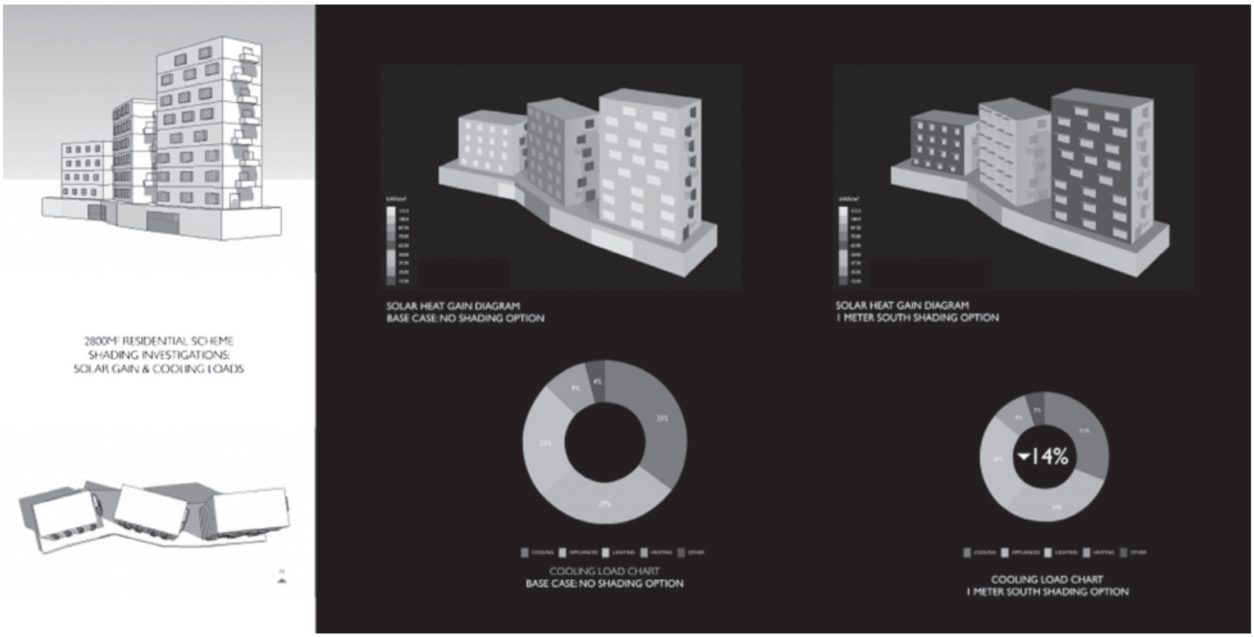


Figure 1.3: Shading tests and corresponding changes to cooling loads. © *Sefaira*

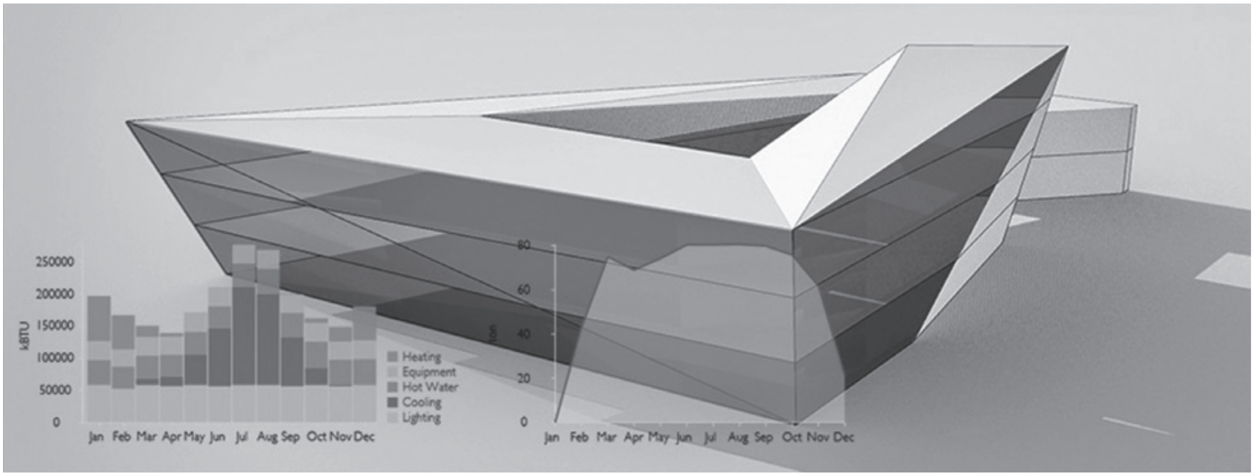


Figure 1.4: Sefaira's outputs include clear informative graphs that can exported and edited to fit the designer's brand. © *Sefaira*

"To us, the cloud is simply a server. There is nothing particularly new about this technology. Architecture firms in the 1970s were using servers for the same reason we use the cloud today: servers can store and process orders of magnitude more data than can be done on a local machine," adds Fano. "Rather than throwing data away, we can keep it in the cloud. We can create massive databases of every model a firm has produced. Not just the final model; we can save every version of the model's development."[3]

"The short answer is that we are really just standing on the shoulders of the phenomenal advances we've seen in computer science in the last three decades," concludes Mads Jensen, CEO of Sefaira. "We live in an incredible age."[4]

Technology is not limited to solutions available to us within our organizations. As Jensen points out, "In many ways, computer games have pioneered models for data-driven decision making. Games like Sim City were way ahead of business software in terms of giving users a data-rich and immersive environment in which to make decisions, and a continuous feedback loop enabling more iterations and ultimately better decisions." (See Figures 1.5 and 1.6.)

People

It's not only about the technology and tools: people make a difference. People are an important force that helps make the gathering, analysis, and application of data a reality today. But not just any people: the *right* people—people with a certain inclination—are helping to make the leveraging of data in AECO industry possible. What these inclinations are vary from person to person, but some patterns can be discerned. The ability to identify and recognize these qualities in others can have implications for human resources, as well as for attracting and retaining talent.

Firm cultures that encourage, or at the very least accept, that working with data is now a significant part of the project team effort can make a difference. In particular, we will need cultures that encourage and uphold the attitudes and mindsets necessary to work with people who are as comfortable working with data and analytics as they are

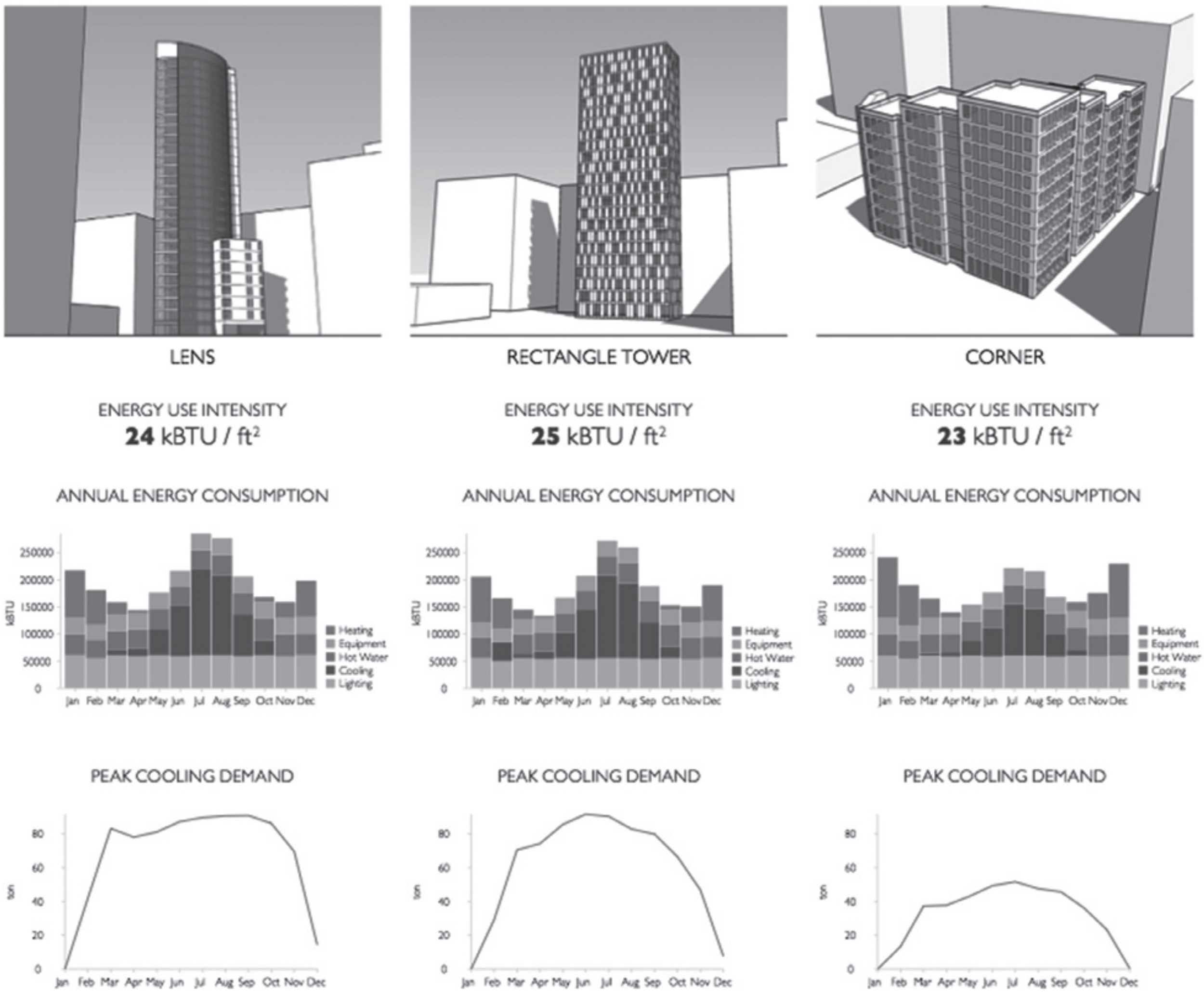


Figure 1.5: Sefaira allows architects to compare design options and measure their performance using chosen parameters. (EUI/ Annual Energy Consumption/ Peak Cooling Demand) © *Sefaira*

putting buildings together. Sean D. Burke, LEED AP, Digital Practice Leader at NBBJ, Seattle, discussed the convergence of parametric and computational tools in terms of people: "From a tools perspective—and tools aren't the only thing causing this convergence—it's the maturity of the design community, everyone being able to take advantage of both ways of working; and a generational thing as well." Due to access to information, ubiquitous training, and the sharing of information, today people are perceived as being more capable of developing the processes and technology necessary to manage data.

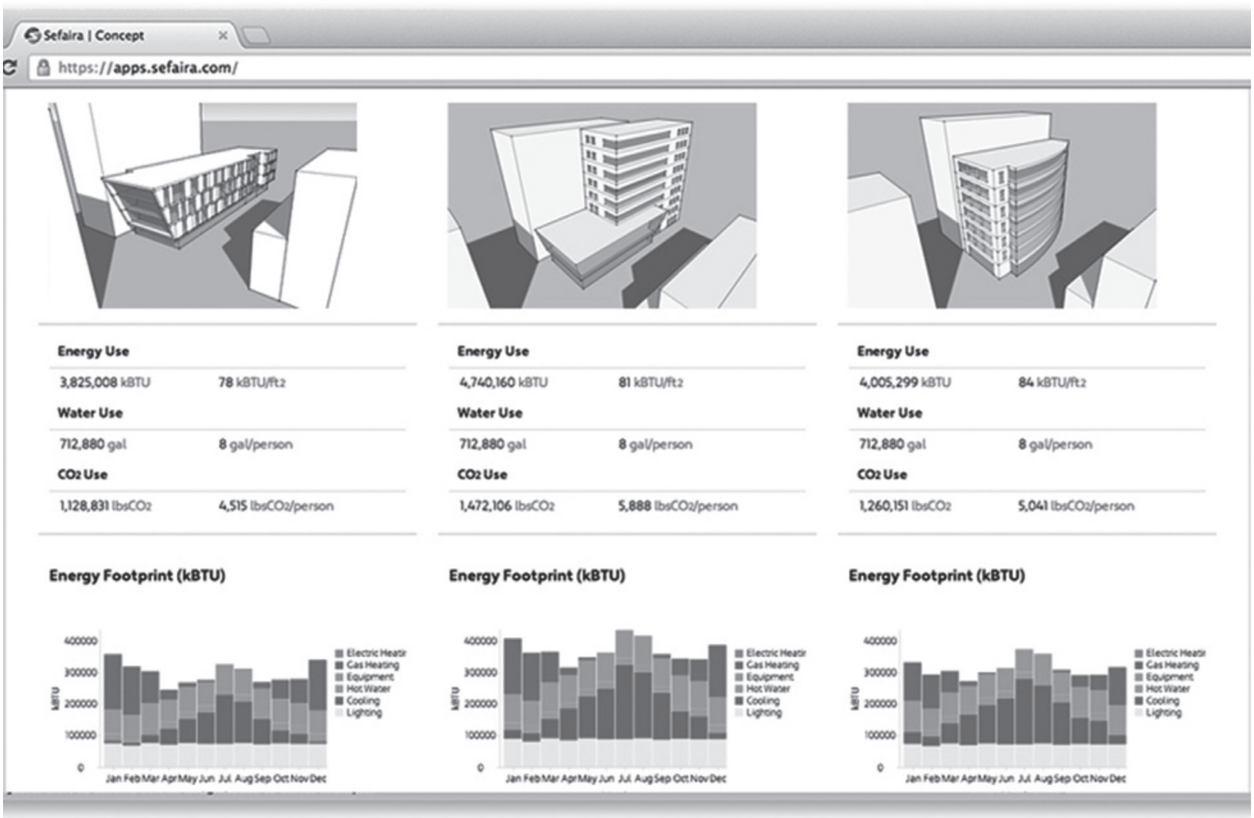


Figure 1.6: Users make comparisons to set the project on the right track early, refine the design as it progresses, and test the effects of design changes (including value engineering). © *Sefaira*

Performance

We're already starting to see a change in the focus of the current generation of architects, from form to performance, away from the media attention of the so-called *starchitect* and creation of monumental, iconic buildings to more site-specific, earth-friendly building interventions. Erik Olsen, PE, Managing Partner, and CEO at Transsolar Climate Engineering, has witnessed the fascination with form taking on a change. "In the younger generation of architects, the fascination with form is not what it was for the older generation of architects practicing today. It's already changing." The move away from an exclusive focus on form has provided an opening for the discovery of data.

Access

There is a lot of data available today, in many formats, and all of it is easier to access today than at any previous time. Although interoperability remains a recurring concern, this is as much due to improved interoperability of software tools as it is to the collaborative, open sharing of information among various parties.

Awareness

Whether through education, enlightenment, or awareness through experience, we are finally coming to accept the nature of the construction industry as being fragmented. In other words, it's an industry built on one-of-a-kind, one-off designs, with geographically dispersed production sites and project stakeholders. Teams come together for a brief time to construct the project, then disperse; notably, these team efforts are marked by a lack of single entities doing it all. The industry is moving from construction being historically risk-averse to assessing and managing risk on a project-by-project basis. Mark Frisch, FAIA, Managing Principal at Solomon Cordwell Buenz, notes that today, "There is a generally greater appreciation of how data can positively inform a variety of processes in our profession." All of these have a role to play in making this time ripe for a data turn. (See Figure 1.7.)

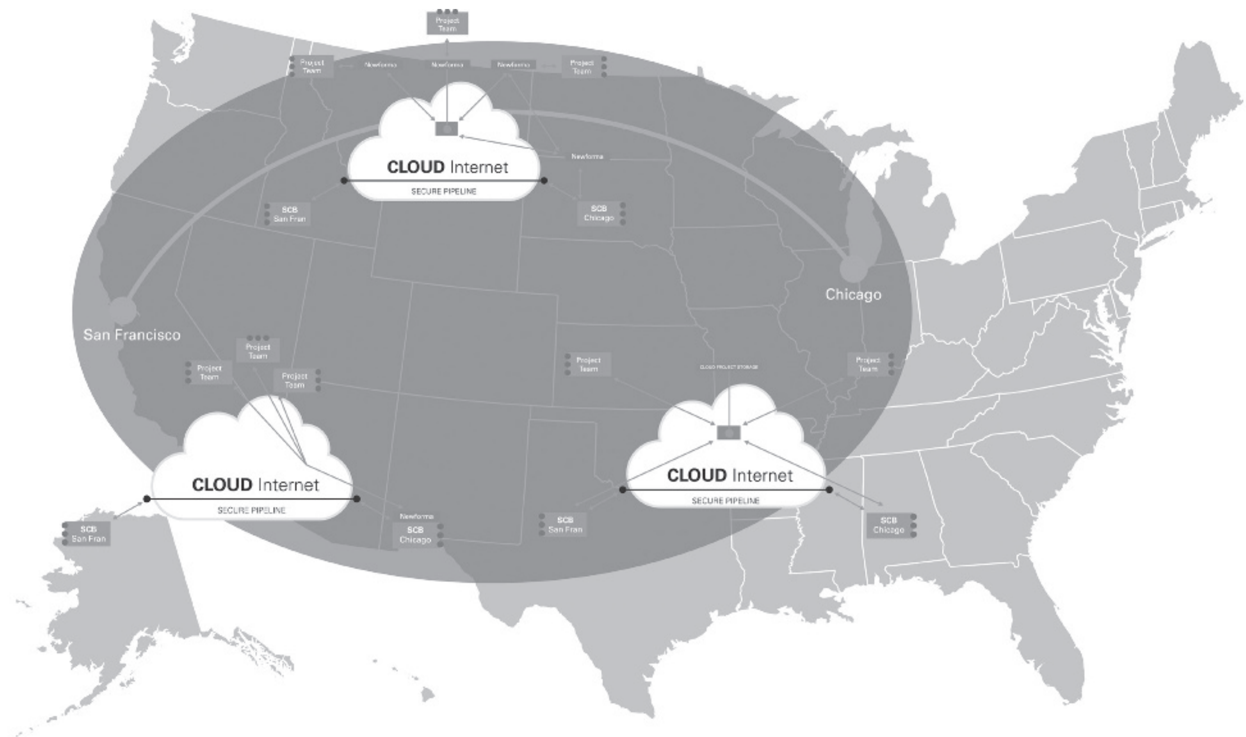


Figure 1.7: Horizon Cloud. Cloud technology enables a secure pipeline for sharing data across offices and project teams. © *Solomon Cordwell Buenz*

Case Study Interview with Robert Yori

Robert Yori is a senior digital design manager at the New York office of Skidmore, Owings and Merrill (SOM), where he co-manages the office's Digital Design efforts and co-leads SOM's firmwide BIM/Digital Design initiatives. He develops BIM curricula for, and teaches at, New York University and elsewhere, and you can often see him presenting at industry conferences, including Autodesk University, ACADIA, and RTC.

Is data something that you just work with and take for granted? Are we potentially fetishizing it by even talking about it?

Robert Yori (RY): *Data* itself is such a broad term. Something that was produced with ink is also data. It's really a question about how it's absorbed, shared, and processed. In the broad sense, dealing with massive amounts of data is something architects have always done, although much of it hasn't historically been computational. The core question is this: How can we utilize the myriad types of data in a way to better our projects? Utilizing computers to do that is also a long and storied history. So the granularity and explicit nature of that data, that's the relatively new part of the challenge. Plus, everyone adopts things at a different pace.

Talk a little about the database work SOM has been doing with CASE Inc.

RY: Some of our recent work echoes an effort we've undertaken in analog form in years past. We've collectivized our knowledge and our expertise about different building types in certain markets. Towers, for example, are a staple of our business. We have a really good sense of the particulars about tower metrics through all of the projects that the firm has completed, through a combination of anecdotal knowledge and rigorous analysis, and we're pretty good about documenting and sharing that. We thought it would be useful to take that to the next level—by moving it out of the analog realm and translating it into the digital realm—for a couple of reasons. One, it allows us some flexibility. It frees the information from the paper documentation that we produce. In paper form, for example, we might decide we're going to publish data internally on eight buildings. If we want to add a ninth, there's a fair amount of effort required to revise the publication, reprint, redistribute, and so on. If the information is part of a database, we're able to flexibly add and remove buildings, markets, and other unforeseen information. More importantly, it enables us to selectively filter in ways that we might not have thought of when we were initially publishing the paper documents. We've been digging deep into analyzing our best projects and putting the results into a database, which becomes a powerful resource for precedent research. [See Figures 1.8 through 1.11.]

> *We could ask the database: "How many buildings have we done in New York with this particular type of glass?" When a client comes to us and says, "I really love that project you did in midtown Manhattan. Can you do something similar for our site in China?" we can begin to analyze our building's key design metrics almost instantaneously, to understand how they may translate to another building in another region. To see whether the glass type is appropriate in terms of solar gain, daylight, or transparency, or R value. To evaluate the cooling and ventilation strategies and determine if they would be applicable in China. It gives us thorough, quick access to a body of knowledge that has historically been difficult to gather at this level of comprehensiveness.*
>
> —Robert Yori, SOM

(Continued)

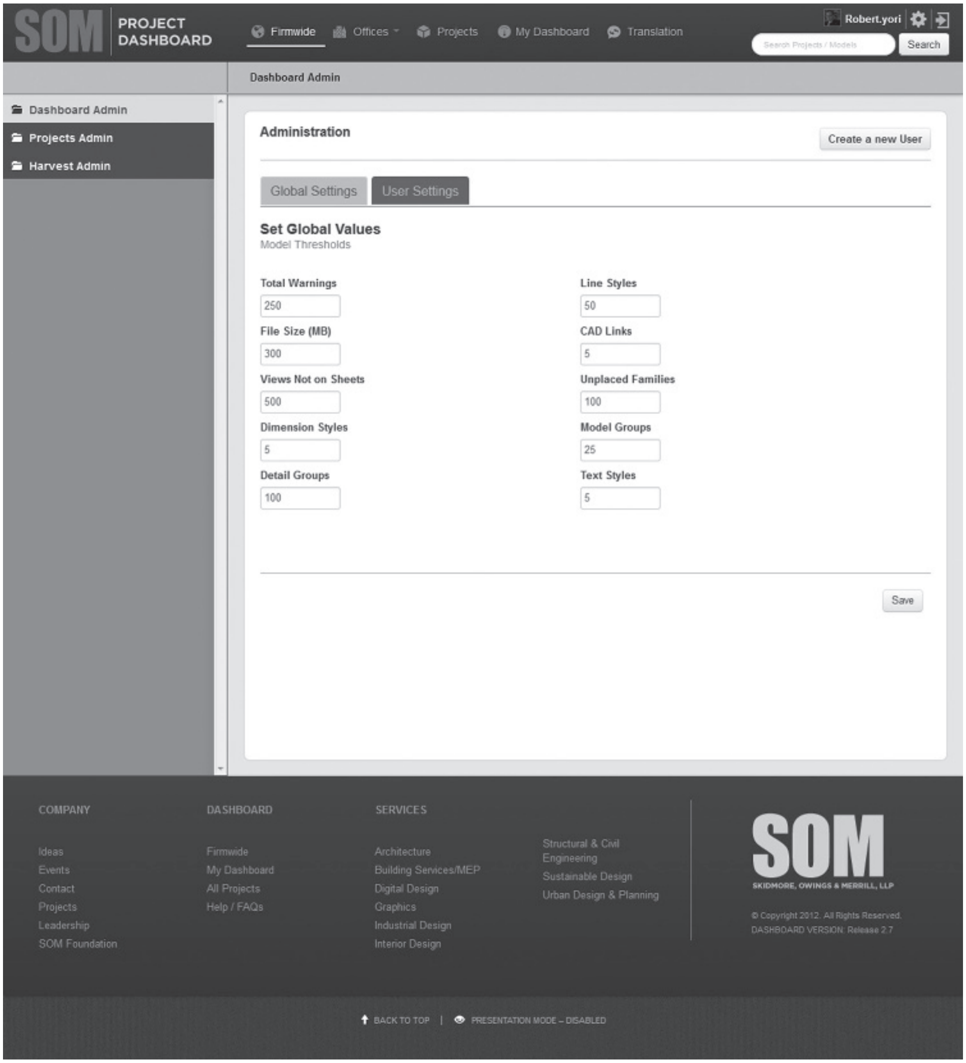


Figure 1.8: The Dashboard can be set to flag properties whose values exceed thresholds set by the user. As the Dashboard grows in functionality, roles can be added or modified. © *Skidmore Owings and Merrill LLP*

How important is it that others in the firm understand that—in addition to their work on buildings and urban spaces—they're also working on databases?

RY: Project leaders and senior architects are juggling vast amounts of data in their heads, and the teams are making it explicit through drawings, specifications, project briefs, and renderings. Over the last decade, as teams have started utilizing Revit, it's been an easier conversation to have. After teams begin to get familiar and comfortable with the tool, I

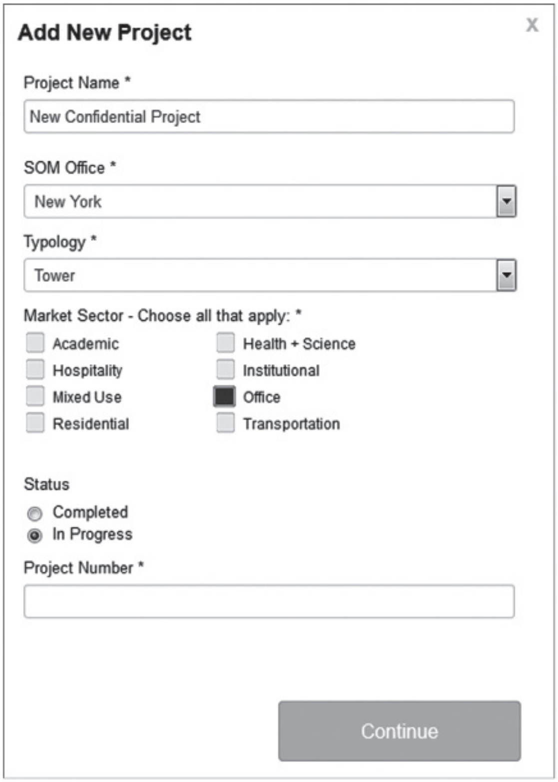


Figure 1.9: Adding a new project involves inputting a number of fields, including market sector, building typology, and status, which can be sourced from an existing database to minimize redundancy and promote data validity. © *Skidmore Owings and Merrill LLP*

say, "You know that's a database you're working in, right?" And many of them respond, "Yeah, I know." It's really a graphic introduction into what a database is and what it might be useful for. Similar potentials existed with CAD, because CAD was a database—if it was used that way.

There are different approaches and varying degrees of understanding and facility with the notion of "drawings as database," just like when computers were first introduced into architecture. As a profession we struggled with the idea of tangibles versus intangibles, what's more difficult to embody digitally, and what can and should be embodied. Overall, we're all having to deal with increasing amounts of data. Those that are computationally inclined naturally would look to some sort of database solution. But I don't necessarily like to call it that from the start—it can scare people off. [See Figures 1.12 through 1.14.]

In the best of all possible worlds, would everyone see buildings in terms of data?

RY: A knee-jerk response would be, "Sure, I wish everybody could do that." I wouldn't be anywhere wise enough to be able to say what the prescription is for the industry. Architecture is manifold. Everybody comes to it with their own

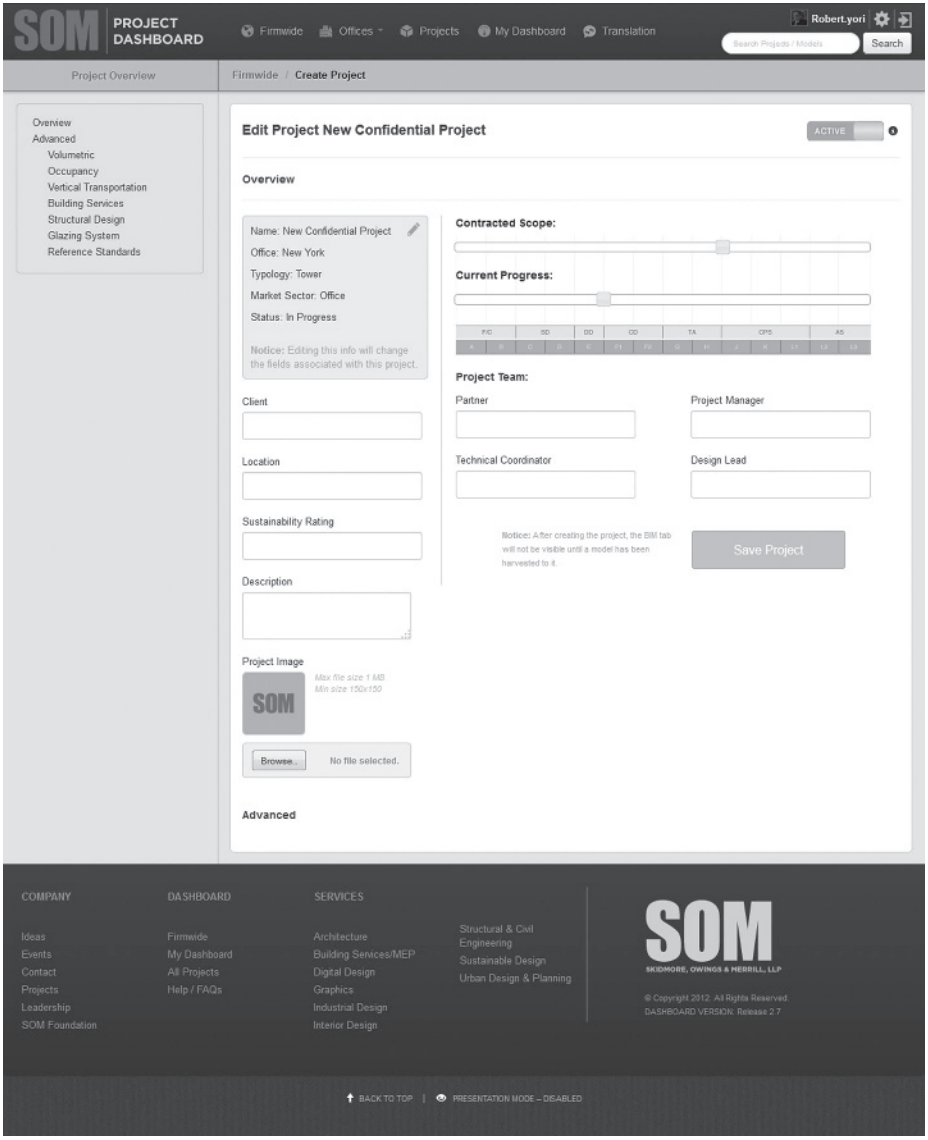
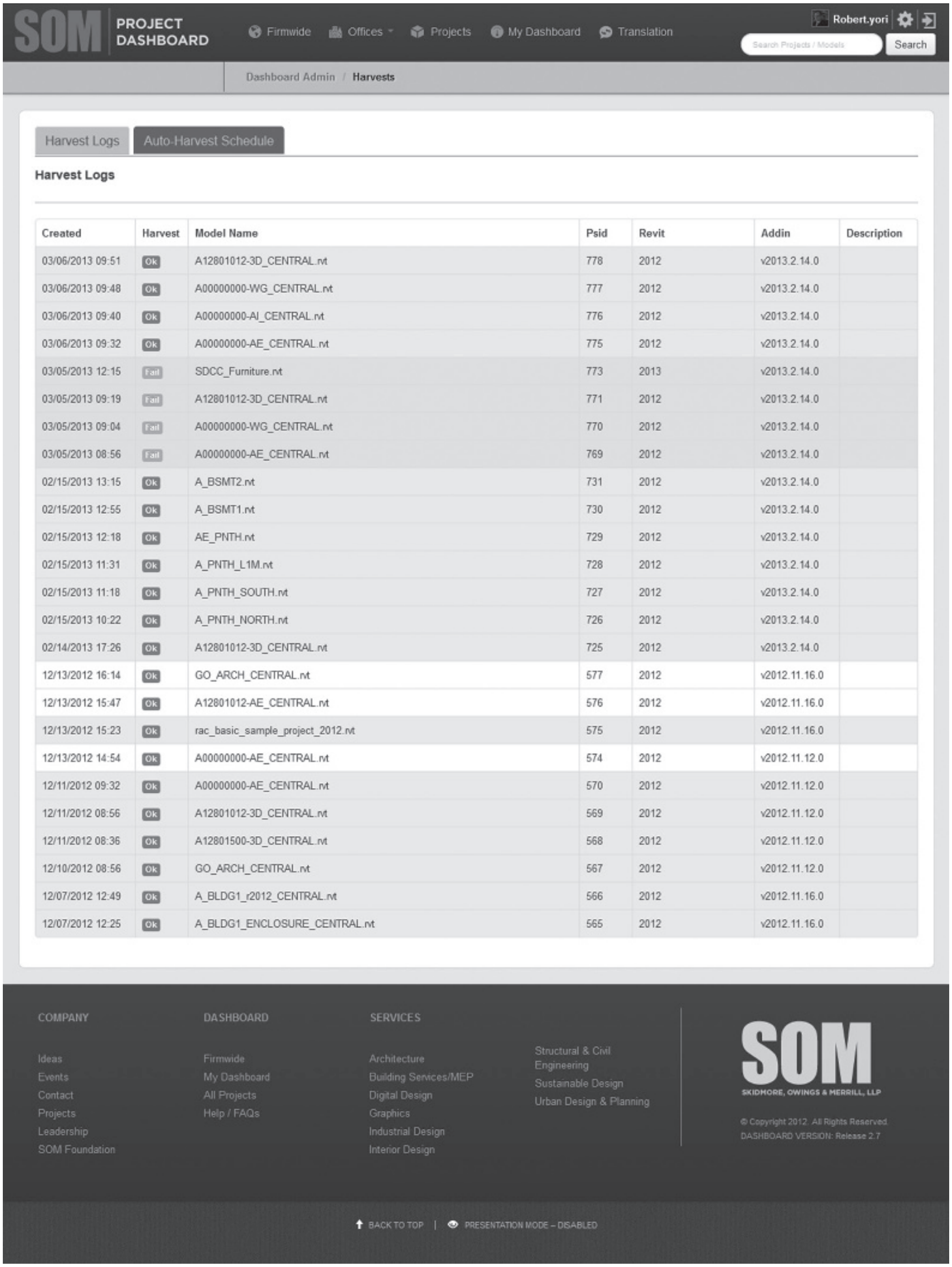


Figure 1.10: Extended information, such as Contracted Scope and Current Progress, can be added for querying projects at a particular phase. © *Skidmore Owings and Merrill LLP*

interests and their own personalities. That's one of the things that make it so fascinating—that it's not simply one set of ideas. As much as I love data, and working informationally, sometimes I'm just drawn to things that are incredibly simple and crafted entirely by hand. It's like the classical music enthusiasts who can't get enough of the three-chord rock-and-roll song. It takes all kinds. So, sure, in some ways it would make our lives easier if everyone would see buildings in terms of data. But I'd be afraid we'd all be missing out if everyone approached it only one way.



Created	Harvest	Model Name	Psid	Revit	Addin	Description
03/06/2013 09:51	Ok	A12801012-3D_CENTRAL.rvt	778	2012	v2013.2.14.0	
03/06/2013 09:48	Ok	A00000000-WG_CENTRAL.rvt	777	2012	v2013.2.14.0	
03/06/2013 09:40	Ok	A00000000-AI_CENTRAL.rvt	776	2012	v2013.2.14.0	
03/06/2013 09:32	Ok	A00000000-AE_CENTRAL.rvt	775	2012	v2013.2.14.0	
03/05/2013 12:15	Fail	SDCC_Furniture.rvt	773	2013	v2013.2.14.0	
03/05/2013 09:19	Fail	A12801012-3D_CENTRAL.rvt	771	2012	v2013.2.14.0	
03/05/2013 09:04	Fail	A00000000-WG_CENTRAL.rvt	770	2012	v2013.2.14.0	
03/05/2013 08:56	Fail	A00000000-AE_CENTRAL.rvt	769	2012	v2013.2.14.0	
02/15/2013 13:15	Ok	A_BSMT2.rvt	731	2012	v2013.2.14.0	
02/15/2013 12:55	Ok	A_BSMT1.rvt	730	2012	v2013.2.14.0	
02/15/2013 12:18	Ok	AE_PNTH.rvt	729	2012	v2013.2.14.0	
02/15/2013 11:31	Ok	A_PNTH_L1M.rvt	728	2012	v2013.2.14.0	
02/15/2013 11:18	Ok	A_PNTH_SOUTH.rvt	727	2012	v2013.2.14.0	
02/15/2013 10:22	Ok	A_PNTH_NORTH.rvt	726	2012	v2013.2.14.0	
02/14/2013 17:26	Ok	A12801012-3D_CENTRAL.rvt	725	2012	v2013.2.14.0	
12/13/2012 16:14	Ok	GO_ARCH_CENTRAL.rvt	577	2012	v2012.11.16.0	
12/13/2012 15:47	Ok	A12801012-AE_CENTRAL.rvt	576	2012	v2012.11.16.0	
12/13/2012 15:23	Ok	rac_basic_sample_project_2012.rvt	575	2012	v2012.11.16.0	
12/13/2012 14:54	Ok	A00000000-AE_CENTRAL.rvt	574	2012	v2012.11.12.0	
12/11/2012 09:32	Ok	A00000000-AE_CENTRAL.rvt	570	2012	v2012.11.12.0	
12/11/2012 08:56	Ok	A12801012-3D_CENTRAL.rvt	569	2012	v2012.11.12.0	
12/11/2012 08:36	Ok	A12801500-3D_CENTRAL.rvt	568	2012	v2012.11.12.0	
12/10/2012 08:56	Ok	GO_ARCH_CENTRAL.rvt	567	2012	v2012.11.12.0	
12/07/2012 12:49	Ok	A_BLDG1_r2012_CENTRAL.rvt	566	2012	v2012.11.16.0	
12/07/2012 12:25	Ok	A_BLDG1_ENCLOSURE_CENTRAL.rvt	565	2012	v2012.11.16.0	

Figure 1.11: Project uploads, or Harvests, can be checked, tracked, and verified. © *Skidmore Owings and Merrill LLP*

(Continued)

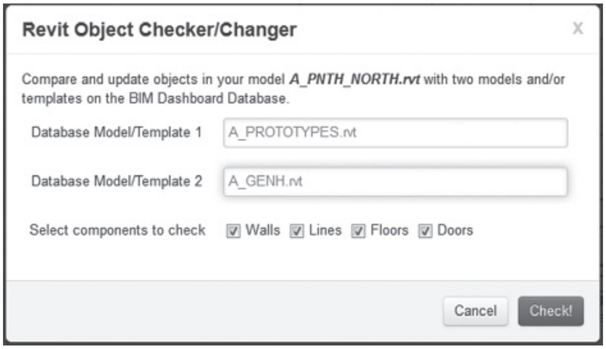
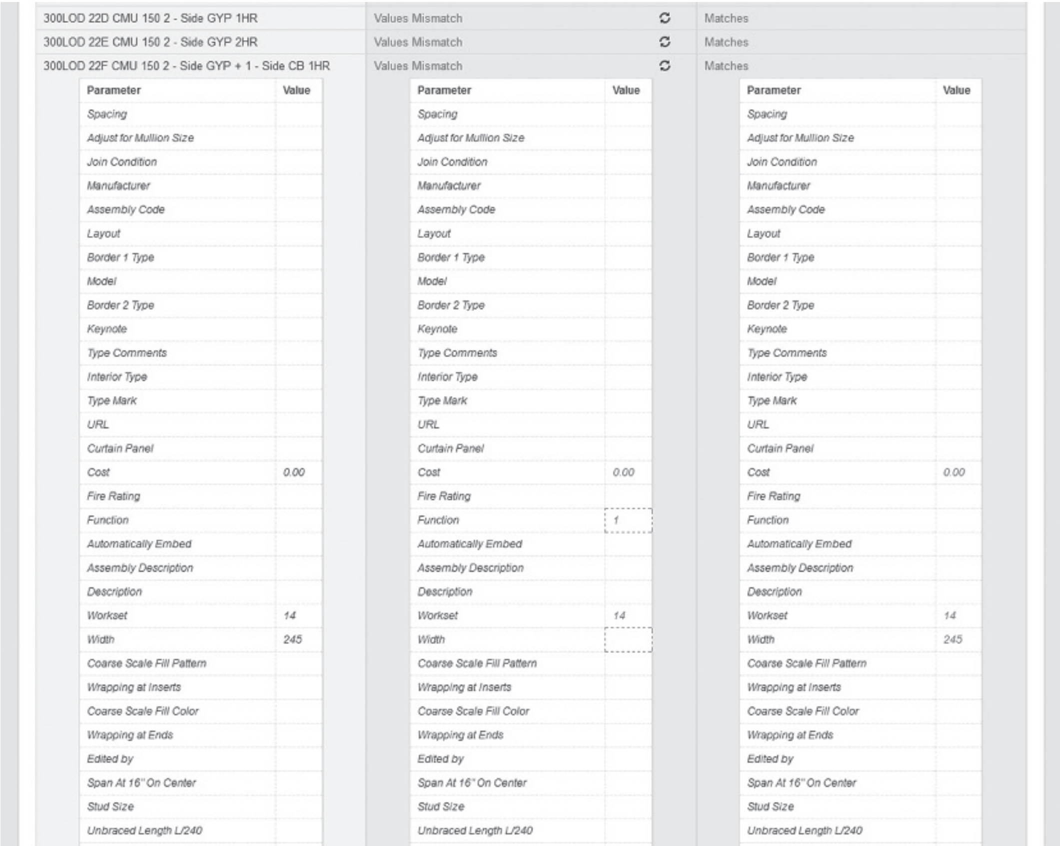


Figure 1.12 A Comparison Engine enables a user to check one or more Family Types against Types in another file, such as a Standards file. Results display discrepancies in Families' Parameter values. © Skidmore Owings and Merrill LLP



300LOD 22D CMU 150 2 - Side GYP 1HR		Values Mismatch		Matches	
300LOD 22E CMU 150 2 - Side GYP 2HR		Values Mismatch		Matches	
300LOD 22F CMU 150 2 - Side GYP + 1 - Side CB 1HR		Values Mismatch		Matches	
Parameter	**Value**	**Parameter**	**Value**	**Parameter**	**Value**
Spacing		*Spacing*		*Spacing*	
Adjust for Mullion Size		*Adjust for Mullion Size*		*Adjust for Mullion Size*	
Join Condition		*Join Condition*		*Join Condition*	
Manufacturer		*Manufacturer*		*Manufacturer*	
Assembly Code		*Assembly Code*		*Assembly Code*	
Layout		*Layout*		*Layout*	
Border 1 Type		*Border 1 Type*		*Border 1 Type*	
Model		*Model*		*Model*	
Border 2 Type		*Border 2 Type*		*Border 2 Type*	
Keynote		*Keynote*		*Keynote*	
Type Comments		*Type Comments*		*Type Comments*	
Interior Type		*Interior Type*		*Interior Type*	
Type Mark		*Type Mark*		*Type Mark*	
URL		*URL*		*URL*	
Curtain Panel		*Curtain Panel*		*Curtain Panel*	
Cost	*0.00*	*Cost*	*0.00*	*Cost*	*0.00*
Fire Rating		*Fire Rating*		*Fire Rating*	
Function		*Function*	*1*	*Function*	
Automatically Embed		*Automatically Embed*		*Automatically Embed*	
Assembly Description		*Assembly Description*		*Assembly Description*	
Description		*Description*		*Description*	
Workset	*14*	*Workset*	*14*	*Workset*	*14*
Width	*245*	*Width*		*Width*	*245*
Coarse Scale Fill Pattern		*Coarse Scale Fill Pattern*		*Coarse Scale Fill Pattern*	
Wrapping at Inserts		*Wrapping at Inserts*		*Wrapping at Inserts*	
Coarse Scale Fill Color		*Coarse Scale Fill Color*		*Coarse Scale Fill Color*	
Wrapping at Ends		*Wrapping at Ends*		*Wrapping at Ends*	
Edited by		*Edited by*		*Edited by*	
Span At 16" On Center		*Span At 16" On Center*		*Span At 16" On Center*	
Stud Size		*Stud Size*		*Stud Size*	
Unbraced Length L/240		*Unbraced Length L/240*		*Unbraced Length L/240*	

Figure 1.13: Results display discrepancies in Families' Parameter values for easy management of multi-model projects. *© Skidmore Owings and Merrill LLP*

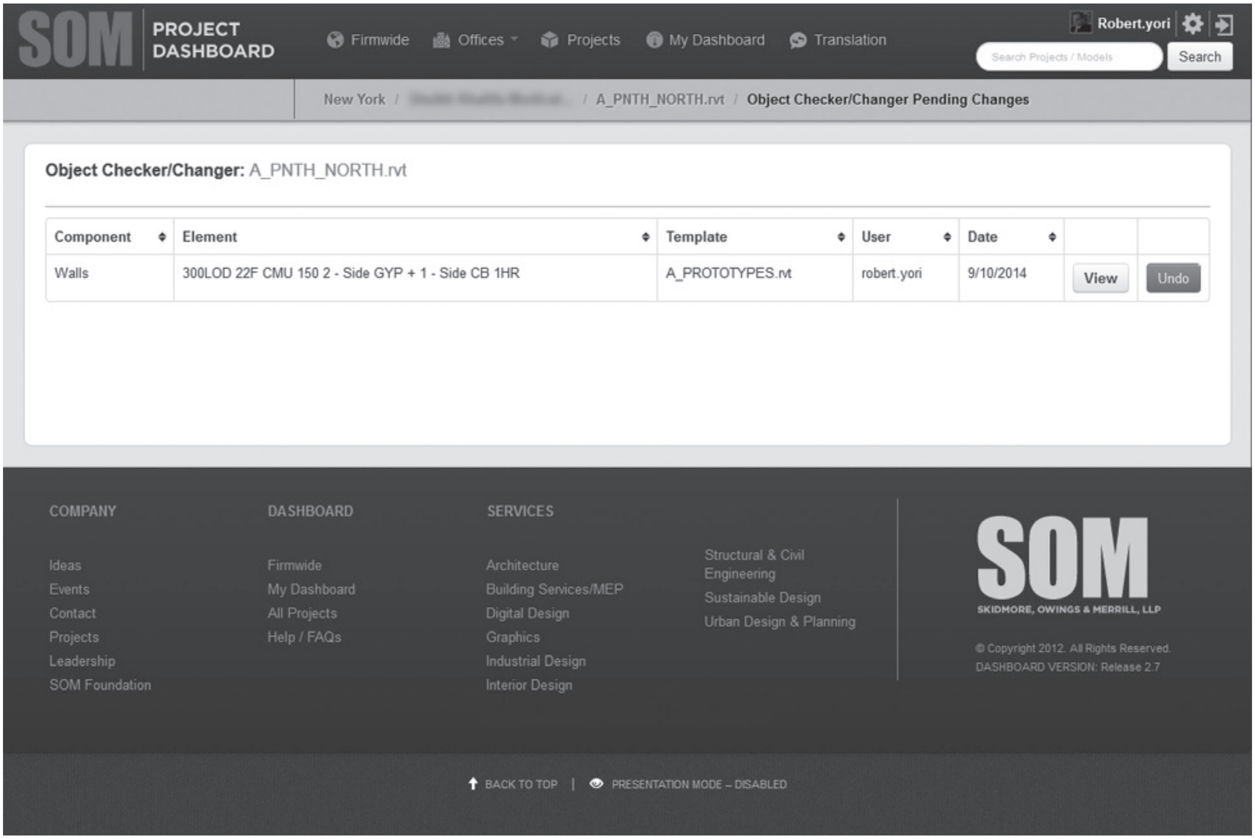


Figure 1.14: Corrections can be made from the Dashboard console and propagated back to their respective models. © *Skidmore Owings and Merrill LLP*

Are there particular technologies that are better at handling project data? Is this ever a factor in your considering working with these tools?

RY: One that is fairly ubiquitous and a great entry point is Excel. How many people create lists in Microsoft Excel? Nearly everybody works in it, many without even realizing that it can be the basis for a database. I've seen that happen a lot—not just here at SOM but all over. I've seen it in my own use. I write down a number of things, then think, well, if I put it in Excel I can do a number of calculations. After a while I say "wouldn't it be great if I could take that and extend it out, and include 'x,' 'y,' and 'z,' and do some calculations, and validate my ideas…" and then the spreadsheet becomes a tremendously useful, ad hoc database. The lowest-threshold, lowest-cost, lowest-hanging fruit that you can do to begin to understand how to utilize what you have is in Excel. Revit is good for this, too, because it provides a database with a graphic front end. For hardcore data gathering, it's a gateway drug of sorts. You see the value of good data, and the possibilities, and begin to look elsewhere for more capable and more sophisticated tools. Oftentimes that requires a more sophisticated or deeper level of knowledge such as SQL databases or more sophisticated modeling programs. But Excel is ubiquitous for so many people that it's a great place to start.

(*Continued*)

Revit is good for this, too, because it provides a database with a graphic front end. For hardcore data gathering, it's a gateway drug of sorts.

—Robert Yori, SOM

Back to CAD for a moment, it has a tremendous capacity to be informational, but, as I mentioned, only if it's used that way. Through my early career I've seen many people not understand the data value of putting things on the proper layer. Or naming blocks properly so they can be counted. One of the complications of any of the computational tools is that their perceived validity can be an all-or-nothing prospect. It has to be perfect. Once someone begins to cast doubt on the legitimacy of the information that's driving it, it can cast doubt on the legitimacy of the entire procedure.

For many design professionals, the subject of data isn't nearly as compelling as the generation of interesting form. Do you see this as an impediment to data use in the AEC industry?

RY: Data is a means to an end. So much of what we do can be classified that way, too. Understanding the motivations behind the data wrangling, and finding value in those motivations, is a conversation that should be had. Putting out data for its own sake might be interesting as part of the process, but it is very much part of the process and not an end goal or solution. People don't generally get into architecture to data wrangle. People get into architecture to solve particular problems or pursue particular interests that they want to pursue. Sure, some are interested in minute problems of great detail—which is great because not everybody is. Understanding data and computational process as a means to an end is really, really important. Because as adults learn, we need a motivation to understand why we should do things differently than we're comfortable with. If we can't find a personally compelling and beneficial reason to change, we won't.

In the near future, what do you see as an ideal firm approach: to strive to be a data-enabled, data-informed, or data-driven practice? What is your firm's approach? Why?

RY: Like any good academic, we should define what each of these three terms means. "Data-enabled" may be being aware of the data but not leveraging it. "Data-informed" might be using data as a factor in the decision-making process. "Data-driven" could imply—I don't know if it's a good thing—that it is your primary priority. I can't characterize the whole firm one way, but certainly aspects of what we do at SOM are data-driven. And some are data-informed. There is some information that is better suited to being data-driven and some that is less so. So holistically, when we are approaching design, I would have to go with data-informed. Because there are some things that we do that are incredibly data-intensive. Some things that we do aren't so much.

I see the ideal approach for the industry as being data-informed, although it is hard to generalize at that level. There are certain types of practice that are more data-driven. For example, my good friend has recently gone to work for a firm that focuses on healthcare. There are lots of fantastic, incredibly fascinating conversations about evidence-based design. A firm doing that kind of intensive work may be closer to data-driven. If you as a client want to go to a more sculptural architect, because you may be looking for something maybe a little less programmatically defined or rigorous, and want something that's more emblematic, perhaps you're closer to data-enabled. Being aware of data and understanding the role it can and should play in one's practice is very, very important. Having an awareness of it. In school, our professors often told us that architecture is about the problems you choose to solve—I would extend that and say "and how we choose to solve them." As long as you are aware of the "data factor," and you're understanding when

When I hear the term "data-inspired," it sounds as if there's an attempt to make it appear as though data was used in an integral way but really wasn't. Is "data-washing" a term yet?

—Robert Yori, SOM

it might make sense to use it in your practice, and to what degree, that's key.

How much of leveraging data is technology and how much is mindset?

RY: You've got to have the mindset first. If you're not motivated to do it you're not going to do it.

What mindsets would you recommend others in our organizations, profession, and industry develop in order to work with data?

RY: If the goal is to get that motivation, I would look to work that is data-driven or data-informed. Dig in and find out how data-driven or data-informed work is improving the quality of the projects and process. And it's got to be fun. You've got to have fun while you're doing it. That's the greatest motivator for anybody.

Can you describe a project where use of data led to an improved decision, insight, or outcome?

RY: We could cite any number of our performance-oriented buildings. So much of that design is data-driven. It has to be. A recent tower project in Guangzhou, China, made significant use of simulation and data-driven analysis in shaping the building to channel high-velocity wind through energy-generating turbines. We're using a similar strategy on a tower in Indonesia that also employs geothermal strategies. We're doing a net-zero energy school in Staten Island, New York, and using data-driven strategies to exceed our goals for solar panel surface area requirements.

Another good example, although not a building, is our Revit and BIM standards initiative. A number of years ago I gave a lecture at Autodesk University on the crossing over from pioneering use to mainstream platform use of BIM, including Revit. I referenced Geoffrey A. Moore's *Crossing the Chasm* very heavily. A lot of that transition involved creating standards for everyone—guidelines, best practices, and so forth. We knew that we had amassed a number of successful projects in Revit done by our pioneers and their teams. And we wanted to figure out how we could triangulate those successes into a body of documentation to be a guide on

Figure 1.15: The BIM Dashboard's front page gives the user an at-a-glance, high-level understanding of norms for file size, project versions, models by discipline, status of most recent and active projects, and more. © *Skidmore Owings and Merrill LLP*

(*Continued*)

Figure 1.16: Users can drill down and visualize specific anomalies. The graph indicates that the largest project has a third more models than the next largest one, and that the majority of projects have one to six models. © *Skidmore Owings and Merrill LLP*

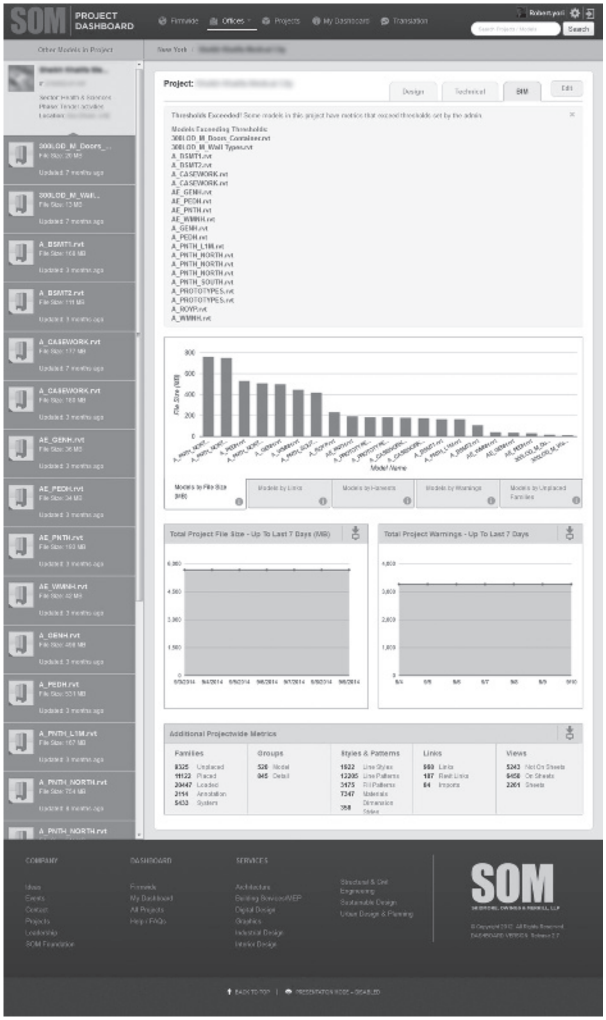


Figure 1.17: The Project page shows information about all of the models that constitute the Federated Model. The file size and aggregate number of warnings for the overall project remain constant—a sign that the project is very well managed. © *Skidmore Owings and Merrill LLP*

Figure 1.18: The Project Model Page gives an immediate status of the health of a project based on a number of commonly agreed-upon metrics. The Model's history is also included, providing additional insight as to future performance. © *Skidmore Owings and Merrill LLP*

(*Continued*)

future projects—in essence, a standards effort. [See Figure 1.15 through 1.18.]

We were faced with two options: Option one, to do what everybody does when it comes to standards. Sit around the table and verbally duke it out over which process is better, what we think this parameter should be named, why we should put it here or there, and so on. Option two was an entirely different approach—to find a way to transform the data, information, and knowledge embodied and embedded in the projects that we had already completed successfully. We chose option two, and began our engagement with data and CASE. We talked to CASE and it seemed like a much better idea to build a tool to query and extract the information out of those models and analyze it.

For example, we looked at our walls. A typical question was how to indicate fire ratings. Should they be described as “1 hr.,” or “1,” or “60,” for minutes? We selected 10 projects that we had all agreed were the most successful ones, harvested the data from them all, and analyzed that data to see how it had been done. It helped us determine a trajectory for moving forward. We weren't blindly guessing. In the Option 1 scenario, the people sitting around that table verbally duking it out, they had that data, but it was only in their heads. It wasn't made explicit and analyzable to the degree that it was in Option 2, when we all analyzed it together. Out of that process we were able to understand amazing things from our projects. Different modeling and logical approaches, naming techniques—everything from the mundane to the sublime. It informed us tremendously in what we should do moving forward.

We got very little resistance because we, as a firm, knew the projects had already worked. This was an evolution of what we had already done and was an attempt to broaden that usage out. This is how we had done it at SOM successfully so far, and used that knowledge to move forward. [See Figures 1.19 thru 1.21.]

Figure 1.19: The Warnings functionality logs each warning from a model and remembers elements associated with that warning, allowing the user to track unique warning instances. © *Skidmore Owings and Merrill LLP*

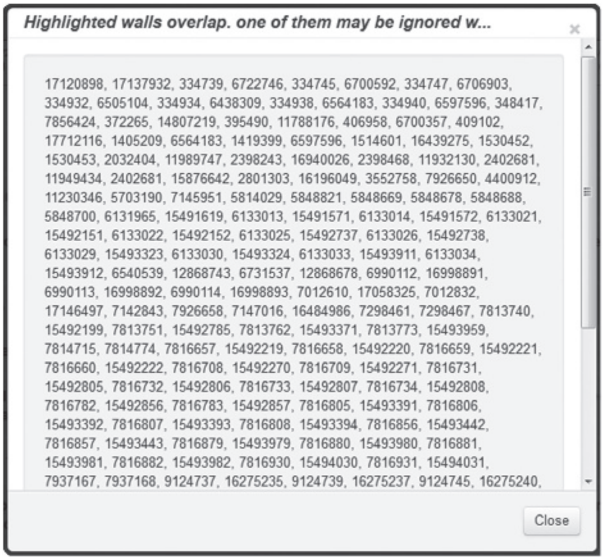


Figure 1.20: A text box provides easy copy/paste access to Warning Element IDs so they can quickly be selected in Revit. © *Skidmore Owings and Merrill LLP*

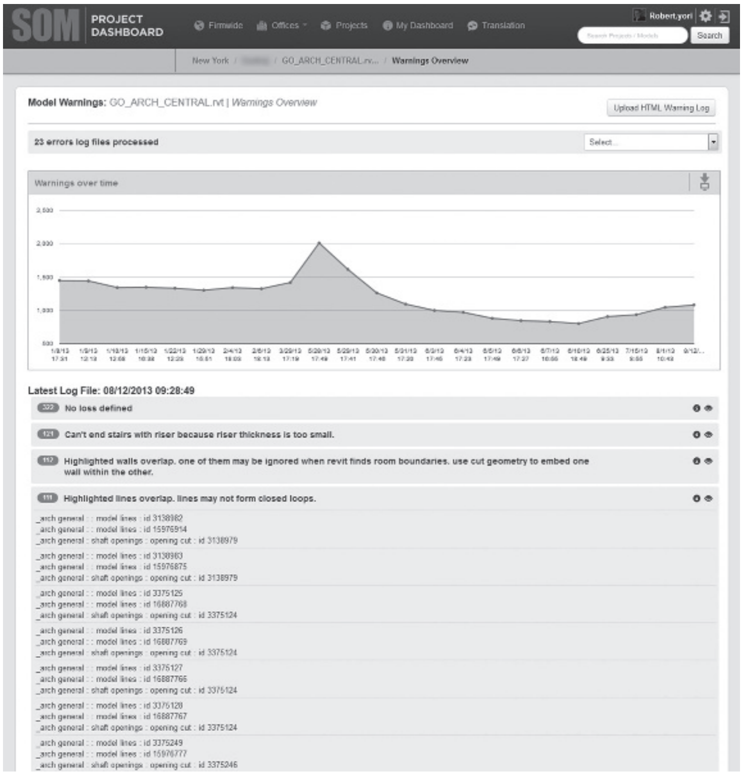


Figure 1.21: Any of the Warnings can be expanded, revealing Element IDs that are indicated below. Each grouping indicates a unique Warning Instance. © *Skidmore Owings and Merrill LLP*

Data Defined

For a book dedicated to the subject of data, when starting out it is important to define our terms. What exactly do we mean when we say *data*? Is it a raw resource? How is it distinguished from information? Is the term so inclusive that it cannot be defined? To define data, we need to look at the various types and quantities of data.

Types of Data

Brian Ringley, Fuse Lab Technology Coordinator at City University of New York and on the Global Design Technology Team at Woods Bagot, clarifies how he approaches and defines data in his work with AEC technology and AEC education. "In my mind, there are three primary categories of data that AEC technologists and professionals deal with, and I define them relative to their relationship to an element of geometry:

Inherent geometrical data, or the data that's intrinsic to the generation of an element of geometry. For example, the inherent data of a nurbs [non-uniform rational basis spline] surface would include data items such as parameter space, boundary edges, and vertices; a guid [globally unique identifier]; control point and degree counts; and analysis data such as measurements of curvature and draft angle, to name a few.

External generative data, or data that is externally sourced from the generation of a piece of geometry for the purpose of affecting said geometry or iteratively generating new geometry. An example of this is the data used to measure insolation of the surface of a building. The actual data items that can be used to measure insolation are things like a Radiance sky model, Radiance material files, an EPW weather file, and definitions of time durations and intervals for measurement. These data items must interact with inherent geometrical data such as surface normal direction and basic object occlusion to compute insolation values, which can then be used to modify existing geometry or generate new geometry. (This is the bulk of what is considered big data.)

Supplemental BIM data, or data for the purposes of building construction and building operation/life cycle that supplement geometrical data, generally produced within spreadsheets or BIM software. IFC data is an excellent example of this, but even simple data items such as the indication of whether or not a wall is structural within the Revit interface would be an example of this. [See Figure 1.22.]

What makes data valuable rather than serving solely as a commodity? The answer may be in the outcomes we seek and how the data is ultimately put to use. Ryan Mullenix, Design Partner at NBBJ, sees the value not in the data itself but in how the data is used. "One of the most intriguing comments I've heard recently, from a San Francisco futurist,[5] is that data is just data. Data doesn't answer a question. Data is just information. Its importance is in how you take that data and use it to address the problem you are trying to solve. That's been a big focus of ours."

DIKW

Is data the same thing as information? How, beyond granularity, can data be distinguished from information? The "I" in BIM, for example, stands for information. How are data and information different? Is it just semantics? Are the two terms interchangeable? "I use data interchangeably with information," admits David Fano of CASE. "It's such a nuanced distinction. When I talk about them, I tend to use them interchangeably." But are the terms truly interchangeable?

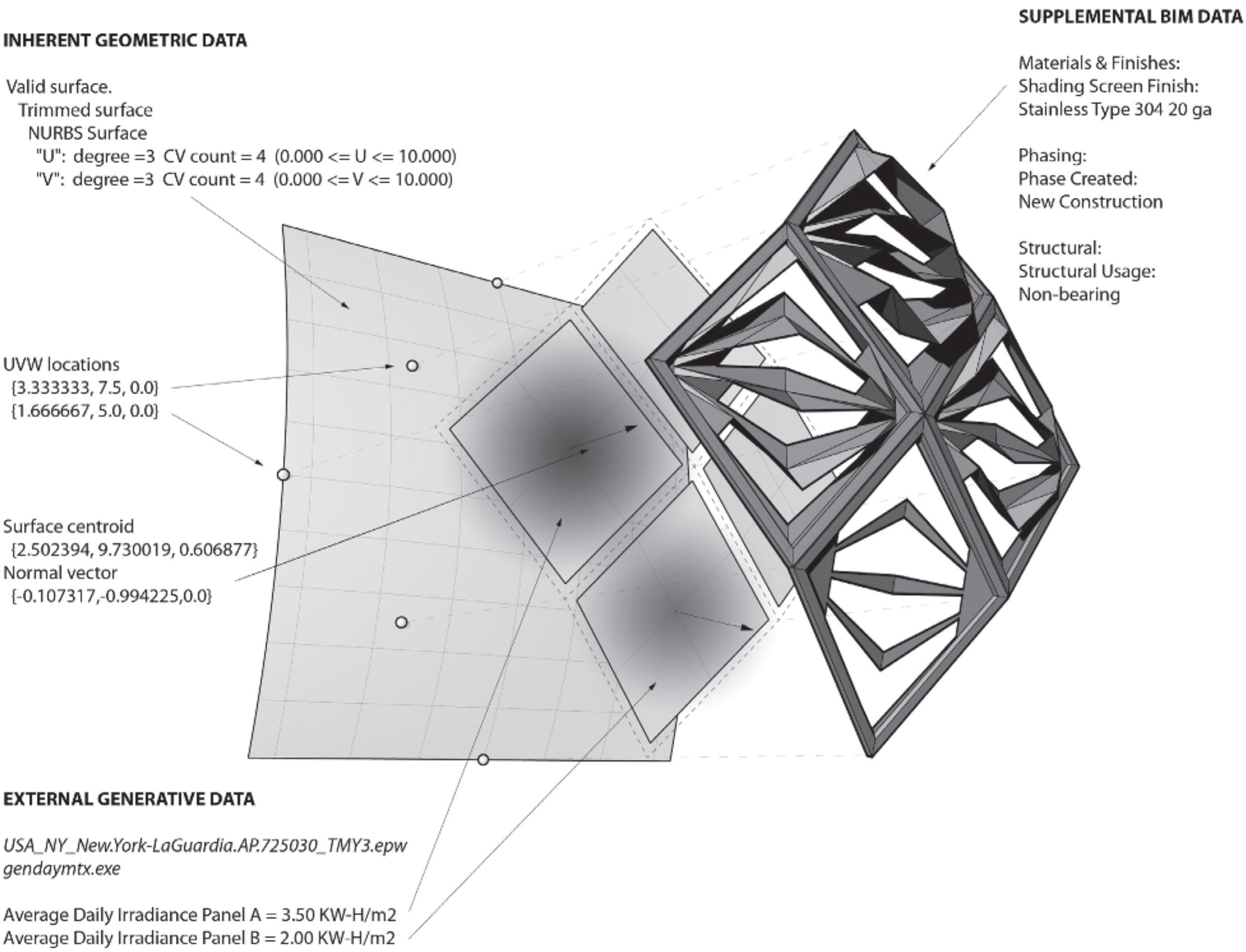


Figure 1.22: Three basic data types in AEC parametric modeling: inherent geometric data, external generative data, and supplemental BIM data. © *Brian Ringley*

We can define data in terms of a continuum, sometimes referred to as the DIKW spectrum or pyramid, where DIKW stands for data, information, knowledge, and wisdom. With the application of data on building projects, "insight" might be substituted for wisdom as a more beneficial goal for leveraging data: data, information, knowledge, and *insight*. Daniel Davis of CASE notes that "most of our industry is based on knowledge and information where we derive insights—whether insights from data or computational tools." See Figure 1.23.

"Conceptually, I believe in the data, information, knowledge, wisdom (DIKW) progression," says Fano. He continues,

> What the industry needs to realize is this is what they've been doing. Part of the reason

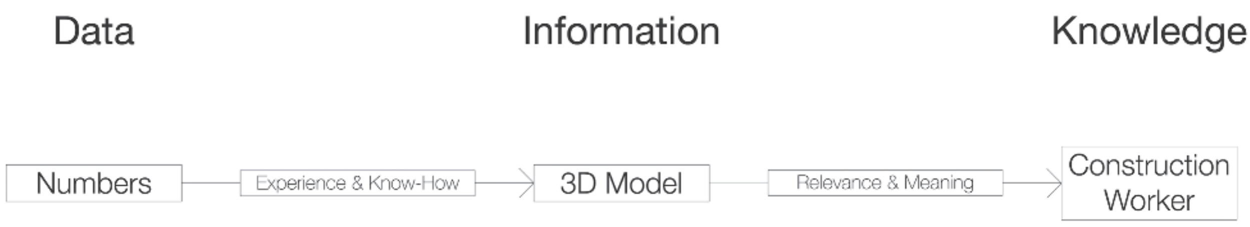


Figure 1.23: DIKW Progression. To arrive at relevant and meaningful decisions, data must first pass through the BIM model. © *R Deutsch*

> architects are so valuable and come into trouble later in their career is because they have accumulated a lot of wisdom. I don't think that can be trivialized. What I think is happening is—if we can capture this stuff which is really only in passive knowledge—now we have all of this more retrievable stuff we can expose the wisdom to a different demographic and one that thinks about things in a different way. I do see this as a watershed moment for the AEC industry. When we could end some of these long-lasting traditions—modes of working—as people begin to leverage information.

Fano describes the DIKW progression in terms of increasing structure to the data: "In its simplest terms, it's using past insight to make future decisions. When it's raw, it's data. When it's a little more structured, it's information. It's about decision making and equipping ourselves with the right things to make better decisions." With the DIKW continuum, it is clear that without data there would be no upstream information, knowledge, or wisdom. Data, in other words, can be thought of as a lower-order or more granular form of information.

One further distinction between data and information can be made. We keep hearing about the "I" in BIM. How is data related to the "I" in BIM? The "I" in BIM is often described as a bookshelf or file cabinet in which manufacturer's manuals and product cut sheets are kept. The BIM is said to *hold* the specifications, the project manual—for safekeeping—so one knows where to find it. In contrast to information, data is at once less specific and more fluid and applicable. With data, the model becomes something more than a receptacle or container where information is stored, more than a retrieval system or long-term storage container. Data in BIM is different in that the data in BIM is fluid and can be queried.

Massive Quantities of Data Defined

As discussed, use of data in the AECO industry is not new. The built environment has long been an abundant source of data. What is new is the amount of data that is available to us; our capacity to measure and ability to capture, process, and act on that data; and, frankly, our industry's urgent need to do so.[6] The use of large quantities of data in decision making in design and construction involves securing a commitment within teams and the organization, reinventing internal and external processes, and modifying organizational behavior.[7] How we refer to massive amounts of data in our industry is still being debated. (See Figure 1.24.)

It is a contention of this book that use of the term *big data*, still popular at the time of publication,

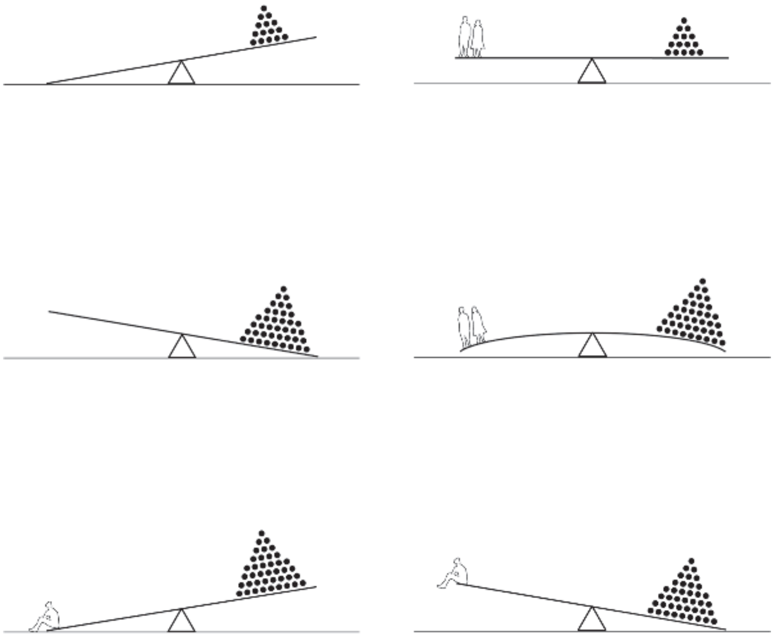


Figure 1.24: Leveraging "big data." Experiment with how your organization will leverage data to make better decisions, bring about better insights, and make better buildings. © *R Deutsch*

will rapidly diminish, and that massive amounts of data will just be referred to as *data*. "Technology is evolving rapidly," acknowledges Mads Jensen, CEO of Sefaira, "and so is the language we use to talk about it. Because of the rapid evolution, we don't always manage to get full consensus on what terms actually mean, before they are either replaced or their meaning morphs again." Jensen took a stab at defining big data: "Big data: Often used as a term for what we can do with statistics once we have lots of data available. We may not understand or be able to model everything that is going on, but there are enough potential relationships that you can start to infer causation and try to draw some conclusions about how things relate."

David Fano finds trying to define big data for the AECO industry as a futile exercise. "If you look at big data, it's just like BIM," says Fano.

> That term came from a marketing department. That term didn't come from anyone actually doing the work. It doesn't matter if it goes away any time soon—it isn't worth wasting energy on. We should embrace it for what it is. It's a mindset. There are definitely ideologies around it that I agree with. So I'll just cherry-pick the ones that work for me to talk about it the way I want to talk about it. To waste any time trying to come to a singular definition I don't see as valuable. When I talk about it, I define it in my terms. I'm going to tell you how I define this term when I talk about it. You can still use your definition. We're all using the same language here. [See Figure 1.25.]

Chris Pyke of USGBC believes that we are getting a little ahead of ourselves with the use of the term *big data*, finding value in the traditional big data

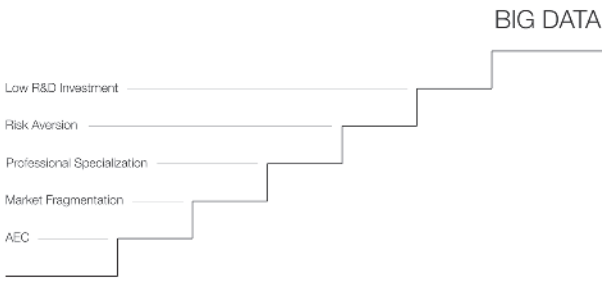


Figure 1.25: AECO industry's considerable challenges to fully participating in big data. © *R Deutsch*

concepts of volume, velocity, variety, and veracity. "Today, our data volumes are relatively modest, velocities relatively low, variety is growing, and veracity is widely (wildly) variable. So, we have some of the elements, but we are hardly approaching big data as it is understood in e-commerce or finance." Pyke continues:

> Big data will come to our industry when we begin to collect and integrate spatially and temporally specific information from millions of buildings associated with billions of occupants using energy and creating social, economic, and environmental impacts on a second-by-second basis. We are creating the foundation for this future, but it remains over the horizon.

Andrew Heumann, leader of NBBJ's Design Computation team, defines massive amounts of data as datasets large enough to require specialized computational infrastructure—such as cloud computing, or farms of machines like supercomputers—in order to process it, and says "under that definition I wouldn't say we're using big data." Heumann goes on to say:

> However, with a slightly more liberal definition, a server with hundreds of BIM models on it is big data—and in that case we use it every day, not just as individuals accessing specific projects, but with our tools that analyze and monitor the performance of all the projects in the firm, taking a look at all the models at once.

Clayton Starr of RTKL defines big data more traditionally, as information gathered to inform the gatherer of trends and to predict future outcomes. "This can be a passive harvest such as my local grocery store loyalty program or a weather station collecting bits of data daily to actively tracking the movement of people and equipment. The biggest surprise is always what you perceive the outcome will be to what it actually is. It can be startling to see how much waste we have in our daily routine, misuse of resources, or how much Kraft Mac and Cheese you actually buy." There are unquestionably fewer specific applications for big data when defined this way.

Strategy No. 4: Not Big Data, Smart Data

Each organization has to define big data in terms that are meaningful for the specific situation and way they intend for its use. For example, Evelyn Lee, a strategist at MKThink, doesn't think about massive amounts of data points solely in terms of size, but rather in terms of what it can do for the client, and says that it's about finding the right balance in everything. Her approach? "We try to pull the smart data from big data." Lee continues,

> Whenever development people say if you want to have the most sustainable building on the block, never turn the lights on. Never run any of the mechanical systems. At the same time, we're trying to produce a productive workplace for your employees. What is the right amount of everything that will get you the highest level of productivity? We do use "big data," and we have a system that can mine it really quickly, but it's really about being smart about the data you're collecting. So we talk about it as smart data.

"Big data companies typically harvest data that is constantly being generated in real time," adds Andrew Witt. "That's never happening on a building project. There may be some collateral information on building projects, but I don't think it can really qualify as big data." Witt doesn't classify his work in terms of working with big data. "When I think of big data, I think of billions of data points. On the projects that we have worked with at GT, they have been more in the hundreds of thousands of data points or maybe millions of data points. In terms of building information, it is really hard to get to that big data threshold with a single project." Witt continues,

> Big data presumes that that sort of data is structurally homogenous and that there's a comparability across all the separate data points. One of the difficulties of talking about big data in the context of BIM is that, taken as a whole, there's a lot of heterogeneous information in the model. All of that information is structurally distinct and it isn't really comparable. You aren't going to compare the metadata of a window with the metadata from a building's concrete slab. They're two different animals. That's one of the challenges when you talk about big data in a context like that. Individual comparable datasets are actually relatively small.

Others contend that big data allows for the comparison of seemingly incompatible datasets. "Look at site selection decisions," suggests Tom Mulhern SVP, Chief Innovation Officer at Dātu Health:

> The real estate data is their data. They're looking at market analyses. They're looking at branch data. At resale value. Their business is built around the mastery of that data. Their ability to process that data on behalf of their client. One of the things that's definitional about big data is overlapping datasets that typically haven't been overlapped. Uniting data about one thing with data about another. Data about the economics of a building overlapped with data about the design of the building.

Case Study Interview with Sean D. Burke

Sean D. Burke, LEED AP, is a senior associate at NBBJ in Seattle, Washington. As the Digital Practice Leader for BIM, Sean is responsible for developing best practices, conducting research and development on new processes and tools, and working closely with the Design Computation group to identify areas where technology can help evolve the practice. Sean has presented at Autodesk University and at conferences around the world.

What implications do some of the new tools have for the sharing of data and even big data?

Sean D. Burke (SB): They're still immature right now. It's hard to say where they're going to go. They're solving an initial niche of peer-to-peer collaboration, in lieu of big, more heavy-handed administration sites that require a lot of IT involvement. I think that's a good thing because it democratizes the idea of project teams. It makes it a lot more agile and reduces the barrier to entry. You can poke into the tool and invite your coworkers and collaborators from other design firms in an ad hoc manner rather than having it be so formalized, where you have to set up an account, give everybody access. It's entirely left up to the individual, which is a good thing. It has disadvantages as well: it's harder to control the flow of information if you have projects that have some sensitivity to the information. The majority of projects, though, don't fall under that category.

(*Continued*)

As for implications for big data, when it comes down to aggregating things across multiple projects or teams, the cloud becomes a pretty rich information source if it can be mined properly, and if access to that data is available in an open way. Currently, the providers of these cloud services, such as Autodesk, are mining that data and creating big data. They might be anonymizing that data and using it for their own internal sales and marketing needs. It's happening already, whether we are benefiting directly or not.

Talk about how big data fits into the BIM workflow. What are some of the ways NBBJ is harnessing big data?

SB: There are a couple different ways. One, we're starting to experiment with ways of getting data out of Revit, and managing it in more of a computer database platform. There are commercial tools out there like dRofus, CodeBook, Trelligence Affinity that are really good on the front end. When you're meeting with a client on a large project like a hospital, and you have to suck in all this data that you have been getting from them, and have to put it somewhere before you've drawn a single line or modeled a single wall. And you want that data to be validated against the model later on once you've built it. Those tools are great for that. We're trying to figure out if there's more opportunity there than those planning tools currently have. We're trying to think about the next step in that area.

On the other hand, we're keeping a real close eye on CASE's Project Dashboard. [See Figures 1.26 and 1.27.] The idea of aggregating data across multiple projects, then putting it in a dashboard-type interface so you can learn several different things, both at the project team level and the business intelligence (BI) level for the firm, is quite interesting.

On the subject of geometry versus data, you've written that[8] "Moving geometry between tools is trivial. Moving data between tools is key." Can you explain how these are different and why the ability to move data is key?

Figure 1.26: The global overview gives a quick snapshot of key statistics that are monitored daily; here the number of active projects and the activity in the BIM models are displayed. © *CASE*

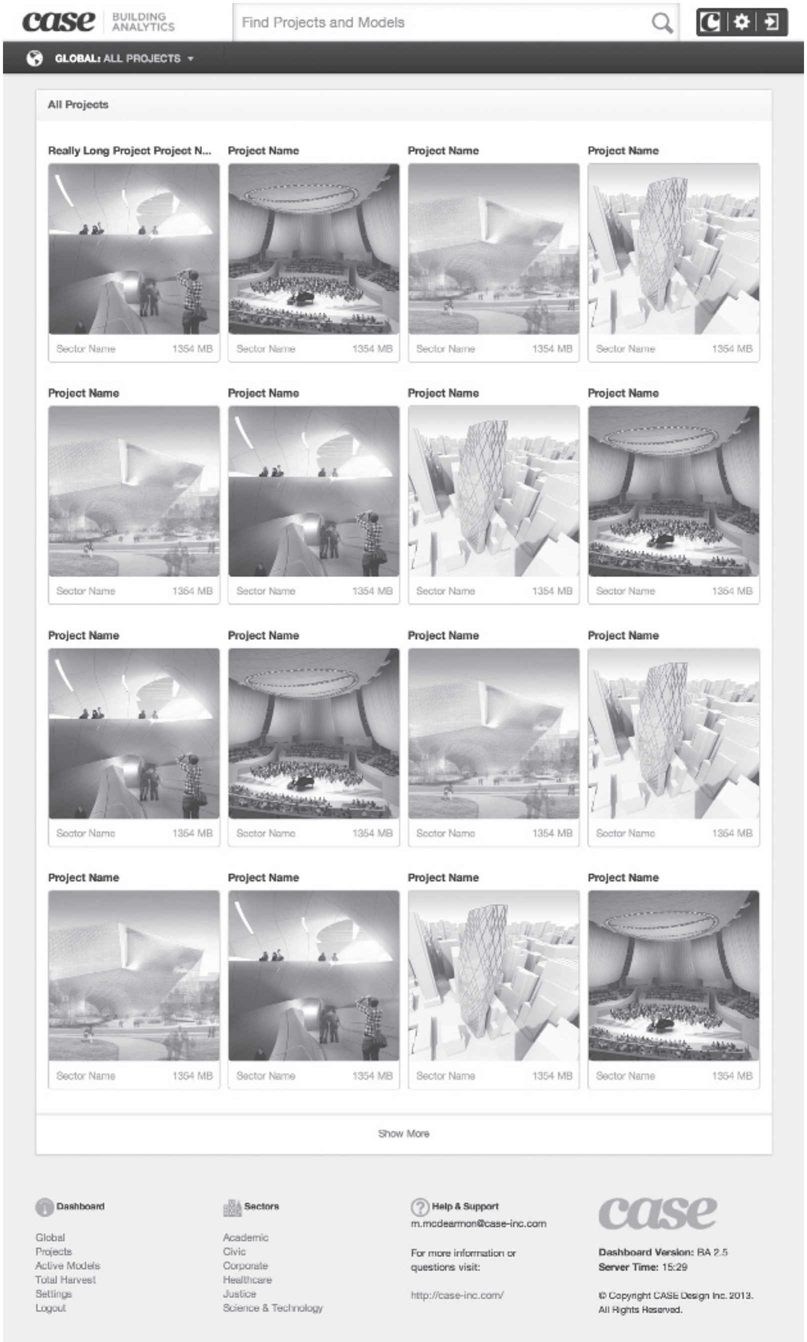


Figure 1.27: The Building Analytics dashboard provides information on every project the firm has done. © *CASE*

(*Continued*)

SB: When data stays in one container for too long, it gets stale. It certainly loses its power. Data, like physical objects, can have momentum. If it sits for too long, it doesn't want to leave. If it's very agile and can be moved from tool to tool, without loss of structure or integrity, that data is much more valuable. Because you can analyze it more easily, append it, or modify it more easily. There's a lot of proprietary software that we use where the information that someone is looking for, like in a Revit model, is there. But the application may not have been designed in such a way that you can access it. A real simple example: floor area ratio (FAR) is a silly, stupid analysis that we should be able to do. But Revit can't compare two different things from two different categories. The building has mass and it has area, it has total floor area, but it knows nothing about the site that it sits on. So a tool like Dynamo can take those two objects and compare that to a formula and say, here's your FAR. You can also very easily hook up some visual feedback as you're designing. You don't need to have someone who is a Dynamo expert use Dynamo as a tool. It can be set up in advance and then minimized. A designer then could be working in Revit and they could be manipulating the massing model. And as soon as it goes out of compliance, it turns red. The whole model just turns red. Then they can push it back down again and it turns green. Just that simple act of connecting two pieces of data, that were already there, in a new way by using another tool is quite a revelation. We think of design computation as something that is about form-making and we're going to have double-curved surfaces. But really it's just a tool. You think of a problem like that where it requires someone manually taking a piece of data and putting it somewhere else. Once data exists in more than one place, it has a tendency of being wrong in both. When data can live in one place as the source of truth, and have connection back to the model, that's a better place to be. If you try to put all your data in one basket versus putting it where it makes the most sense.

A designer could be working in Revit and they could be manipulating the massing model. And as soon as it goes out of compliance, it turns red. The whole model just turns red. Then they can push it back down again and it turns green. Just that simple act of connecting two pieces of data, that were already there, in a new way by using another tool is quite a revelation.

—Sean D. Burke, NBBJ

What needs to be in place for this to happen?

SB: It could be an off-the-shelf tool. For us to be more successful in extending our capabilities and the reach of BIM, we need a little bit of a shift on the part of developers to give us direct access to our data, so we're able to query a Revit model from an external source. Data and geometry—the distinction is so fine. It's still data—it's just graphic data instead of non-graphic data. [See Figure 1.28.] They're both important. The computer doesn't care what's what. We just conceptually separate those two things because our profession is visually oriented. We can't see beauty in the Matrix. Most of us anyway.

The raw data behind the Revit user interface has a lot of secrets to reveal still. We just have to figure out ways we can get at it more quickly and easily. Maybe the file format needs to become open. Maybe its competitors need to take IFC more seriously and build an authoring tool on top of IFC so that there's no translation whatsoever. It's just there in an open schema that anyone can access from any tool. You just take the parts that you need and work at those.

People are getting the wrong impression where Revit's value lies. It's a database. We really need to start treating it like one.

—Sean D. Burke, NBBJ

Does style over substance present a danger in the development of thoughtful architects-in-the-making? Similar to algorithms for geometry versus for building performance. How do you anticipate data will fare?

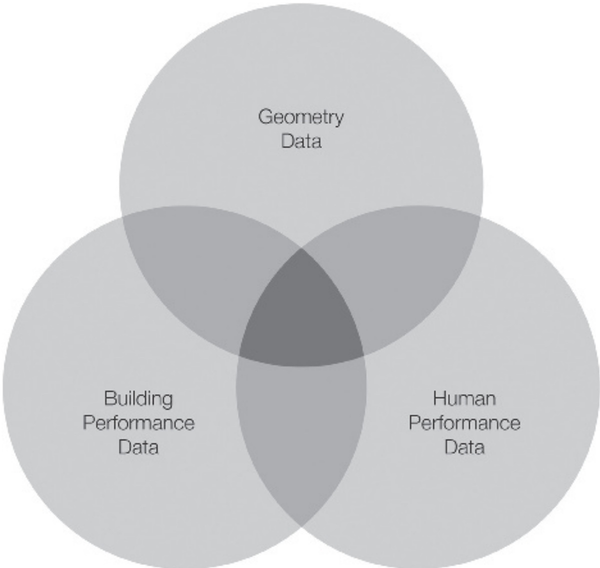


Figure 1.28: Whether geometry, building performance, or human performance, it is all data. © *R Deutsch*

SB: Data will win when it is able to be validated. A concern of mine is how rapidly computation is expanding. I'm part of that expansion. I'm jumping into it, head first, because it has more potential value than BIM alone in how we work. The danger of that rapid expansion is people going in and grabbing algorithms from untrustworthy or unknown sources, putting them into their work, producing a result, showing the client and hanging your hat on that. It could have a severe backlash if we're not careful. I call it the snake oil salesman dilemma. You're standing up in front of the crowd with a flashy presentation, with all of these great graphics. If at the end of the day you give the wrong piece of data, or a piece of data that's interpreted in the wrong way to the client, and they latch on to only that—and that is wrong—the whole thing unravels.

There are two schools of thought when it comes to energy analysis in the industry. One, you're picking a baseline design and you're making it better or worse. It's like going to the eye doctor, and they flip the lenses: this one, or this one? You pick which one seems better. Here's the base, and out of the five different design studies we did, one in five [was] up to 30 percent better than the base. You're basing your decision on relative data. The other school of thought is hitting this exact number. Because that's what the software tells us. You're in early schematic design. You haven't thought of all the factors. You haven't thought of operations or occupancy. There are too many unknowns.

Take two presentations that are otherwise identical: one shows a number, while the other one shows a percentage, plus/minus. I would err on the side of loose interpretation of the results, rather than staking everything on the piece of software that generated it, whether commercial or an open-source algorithm; or the skill of the person who's driving this tool.

I've seen something that was presented that seemed totally out of whack with reality. Diving a little deeper, [I thought] oh, well, OK. This person had never done this analysis, or used this particular tool, before. We have to be really careful. The leadership at NBBJ is keenly aware of these things. And has done a lot of good due diligence with project teams to make sure they understand these risks. It's great to have their buy-in.

What about the sole proprietor? Or the small firm that wants to take advantage of all of these tools and methodologies, but they don't necessarily have the expertise? There's a lot of false confidence that can be gained from seeing a pretty graph that comes from a tool. When later examined, even commercial software can be completely unreliable.

You help facilitate change and transition in dynamic workplaces. Not everybody is comfortable with change. Technology is precarious. What do you advise?

SB: Pick something that you really love. Or something that aligns with your core values. And make that your profession. If your heart's not in it, it's a job. I get really pumped up every time I go to these industry conferences, not because of some new

(*Continued*)

feature that's available, but because I get to talk to all these other like-minded people that really have their heart in it and believe so strongly that this is meaningful work. When I was doing door schedules in AutoCAD that was not meaningful work. Someone could triple my salary and I would never work in 2D again. I want to create value, not suck it from the room.

Someone could triple my salary and I would never work in 2D again. I want to create value, not suck it from the room.

—Sean D. Burke, NBBJ

Back in 2011, you were one of the first people to describe real-time analysis working in a BIM environment. Can you describe real-time analysis—from a data standpoint?

SB: That's not even really real time. That's near real time. This isn't really energy analysis. This is just getting climate data, which to download from the NREL website is very painful, to make it useful. You then have to convert it into some file format that your energy modeling tool can read.

It's thirty times faster than traditional energy modeling because you're actually using your design model to generate the energy model tool. Then it's processed online, off of your computer, so you can continue working. If at the time you ran an Ecotect simulation or TRACE™ simulation, those things can take a long time to set up. And when they're running, it's occupying all of your computer's resources. You can't do anything else. You press a button, then walk away, because your machine is now useless until it's done.

Removing that from the equation is very liberating. You can do a lot more work while it's happening. You don't have to be as selective as you used to be about energy modeling. You never made changes, because you'd have to build a new model used by the engineer. But now, they can take the design model that's been processed in Green Building Studio and convert it back to gbXML, and brought into their energy model, enhanced with more intelligent data. Engineers that are able to work more closely with the architect are embracing this, and are a lot more successful at finding innovative solutions. Start with an optimized building design and add an optimized system design to complement that versus firms that aren't doing any energy analysis. They might be siting their building wrong, creating solar gain because it's facing slightly the wrong way. Using a poor design, then throwing it at the engineer, which is not collaborative. And you're saying, "make it better. Make it meet the minimum requirements." It's nearly impossible to meet the AIA 2030 Challenge by working that way. It has to be more collaborative. The systems integration folks, not necessarily the engineers—sometimes they're one and the same—are going to be better at this.

Moving from near real-time to actually real-time feedback on our work is very near. Our software can do it and our hardware can do it. It's just a matter of the vendors mobilizing to get all of that stuff created as a product and put in our hands. Autodesk may very quickly be challenged by some competitors in this space. There's Sefaira that's pretty close to real-time energy modeling. You're not working in a BIM world, you're still in this loose modeling tool. How do you transition from that to intelligent design data? When you have intelligent analysis data on top of a model you can't use in Revit? [See Figure 1.29.]

Building a BIM tool on top of a modeler is going to have the same challenges that building a BIM tool on top of a 2D CAD solution [had]. Revit's competitors—they're all mired in the fact that they are trying to be a general-purpose platform that has architectural tools on top. They're BIM. They're BIM tools. But they're not a database. And they're certainly not purpose built. Because in AutoCAD Architecture you can explode a wall, and now it's no longer BIM, is it? Just 3D faces and space that have no data attached to them whatsoever. You shouldn't have things that are that easy to cheat. Any editing should be nondestructive. Sure, SketchUp has the capability of creating BIM data. But you have to be so disciplined in how you do it.

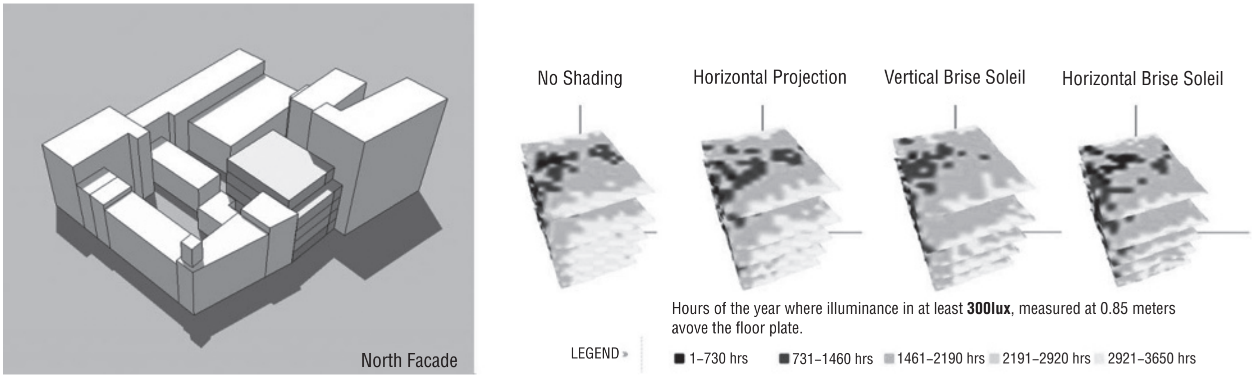


Figure 1.29: Shading analysis using Sefaira's Daylighting Visualization. © *Sefaira*

Data versus Documents

Architects, of course, don't produce buildings. Unless they are working direct-to-fabrication, they produce instructions, in the form of design intent documents, for the making of buildings by others. This is an important distinction lost on many who have never worked with, or as, an architect. Architects have historically associated their value with the production of these documents, whether linen, paper, Mylar, vellum, or digital. As with documents, there are also many sources of data—sensors, BIM models, card swipes, barcode readings, and GPS, to name just a few—just as we have seen that there are many types of data—photos, video, and paper documents among them (these and other types of data are covered in Chapter 4).

But what about documents? Can't documents also be considered data? Or does everything have to become either digitized or datavized to become data? Perhaps the greatest leap forward in recent years has been our turn from being a document-centric industry to being a data-centric one. “Everything is data,” says David Fano of CASE. “Our gripe is not with documents or with paper. Paper's fine. Paper serves a very valuable service.” Fano gives an example:

> Say that a 24 × 36 or 36 × 48 sheet size is the only way building information is conveyed. Why? That's an old thing that came from modes of production at that time. We have iPads now. We have laser printers that can go on the jobsite. Why shouldn't a drawing set be the size of a book? We can zoom in and zoom out now. Scale had to do with the size of a pencil and how much information you can put on paper. We need to recognize the opportunities that current mediums allow for.

“Documents are fine. If you look at the latest trends in databases, they're document-based databases rather than table or relational databases,” adds Fano, and continues:

> What we want to challenge is the presentation of the information. A lot of the thinking in the industry has been about CYA, document it so you can go back and say you did. If it's about giving the right amount of information to the right people at the right time, then we can challenge what all the principles are for what a drawing set is: the documents that are required to build a building. A document for me is a video file. Let's use video. Let's not confine ourselves to 2D abstraction.

Tyler Goss of CASE discusses the movement from architects producing documents to architects leveraging data and the implications for practice and education: "There's a fundamental shift from a document-centric to data-centric delivery methodology in our industry. With a few exceptions, the schools are not preparing people for this. That said, more and more graduates leave school with in-depth practical knowledge of Grasshopper, a parameter-based, rules-based design process. But that shift from a document-centric to data-centric approach, being the one who can lead a practice into making that shift themselves, is going to put themselves in a position of power more quickly than they would otherwise." (See Figure 1.30.)

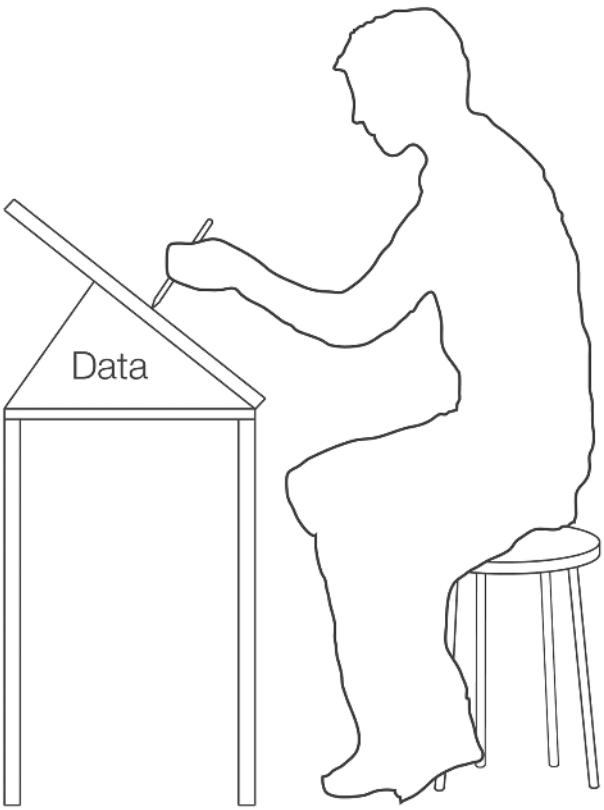


Figure 1.30: Database: Ideas backed up with data is still why many people choose to work with architects. © *R Deutsch*

Goss provides an example of what he means by document-centric thinking in terms that will be familiar to anyone who has worked with Revit and BIM: "Revit can be used in one of two ways. It can be used to build a fundamental logic of a project. In terms of a logic of building. Or it can be used to expediently generate 2D documentation for contractual purposes. More often than not, it's the latter way that Revit is used."

Robert Yori cautions that there are different approaches and varying degrees of understanding and facility with the notion of "drawings as database." He compares the shift toward becoming data-centric with the fearful time when computers were first introduced into the architecture profession. "As a profession we struggled with the idea of tangibles versus intangibles, what's more difficult to embody digitally, and what can and should be embodied," says Yori. "Overall, we're all having to deal with increasing amounts of data. Those that are computationally inclined naturally would look to some sort of database solution. But I don't necessarily like to call it that from the start—it can scare people off."

To help explain this concept, the architect's instruments of service, the building documents, can be compared with data visualizations. "If you look at the rest of the world, data visualization has become this very powerful thing," says Fano. "The *New York Times* will spend a lot of money on the top data visualizer in the world because now you can understand very complex things in a very simple way. So for me, a drawing set is a data visualization. And it is time for that data visualization to evolve."

Architects have stacks of drawings—much of them archived. Should they consider this to be data they can access and use? Mani Golparvar-Fard, PhD, Assistant Professor of Civil and Environmental Engineering and Computer Science, University of Illinois at Urbana-Champaign, thinks so. "Yes, definitely. We can leverage [our proprietary] Mobile Augmented Reality System (MARS) platform to provide near real-time access to the PDFs of these drawings. We can use the interface to perform mark-ups." See Figure 1.31.

Figure 1.31: MARS web-based platform for crowd-sourcing construction activity analysis. Users provide annotations on the role, activities, and tools used by the craft workers and the platform extrapolates information to the video frames. © *Mani Golparvar-Fard Ph.D.*

The distinction between documents and data may soon become moot, due to the advent of BIM where conventional building plans, elevations, and sections can be seen as views of the model database. Zigmund Rubel of Aditazz speaks to this point when he says, "The documentation is an output of the (BIM) model. The model is what's going to get built." He continues,

> In our world today, documentation is what drives what gets built. What we're aiming for is we're going to virtually build whatever's going to get built. The documentation is just to support the regulatory and other aspects of the construction process. The data is what is actually getting built. Documentation is just a report from that. So it's a very different mindset than what is currently considered.

Case Study Interview with Jonatan Schumacher

Jonatan Schumacher is the Director of CORE studio, Thornton Tomasetti's firm-wide, virtual incubator of ideas, where he oversees research initiatives and strategic software development related to workflow automation for integrated building design, analysis, and fabrication methods. Having studied in the fields of product design, architecture, manufacturing, robotics, engineering, and computer science, Jonatan's versatile expertise includes digital fabrication, automatic model creation based on performance parameters, computational analysis, web development, and BIM workflow integration through custom automation. Jonatan lectures and consults on programming, interoperability, and parametric modeling at Stevens Institute of Technology, Columbia University, and the New York City Tech College.

You are the rare design professional who appears to be equally comfortable generating form and optimizing building performance.

Jonatan Schumacher (JS): It is very hard to find good people who are interested in, and able to do, both. There is Mostapha Roudsari, Integrations Applications Developer at Thornton Tomasetti (TT). He's one of those very rare individuals, an architect by training, who is focused on sustainability services and energy analysis. He develops Grasshopper plug-ins for weather data, daylight, and energy simulation. Given his design background, he understands what is important to firms, the process and method of analysis. I knew him originally from the (online) Grasshopper community. He is also very big on Twitter. It's funny how there is this second world where you see people you don't necessarily see at conferences. [See Figure 1.32.]

It is very hard to find a person who can understand automation but also the subject matter. Sometimes we think we should just hire computer scientists. Obviously, we can't pay them what Google pays them. But get somebody who

(*Continued*)

would otherwise work at Google. We had an intern last year who had two computer science degrees. It was very hard to work with him. He was so far removed from the reality that we are still dealing with paper and drawings—boring stuff. It didn't make any sense to him, coming from a different industry. But it is unfortunately the reality. There needs to be somebody who can at least understand how things are done here. Teaching concepts of computer science to architects and engineers helps us.

Everybody is at a point where algorithms are good at automating geometric model generation. That's one thing. The bigger thing is the data that comes with it.

—Jonatan Schumacher, Thornton Tomasetti

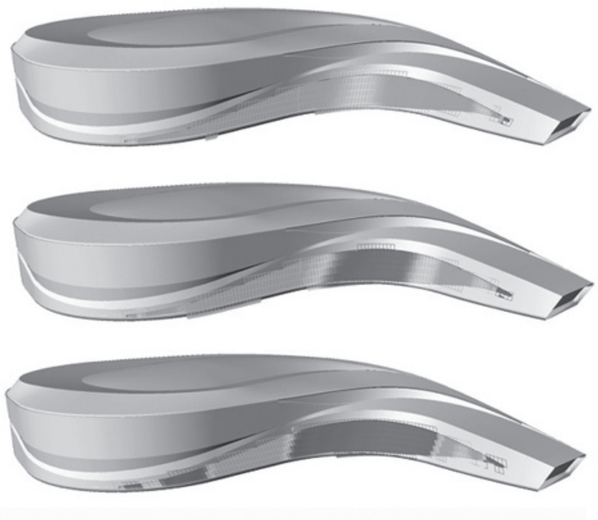


Figure 1.32: Thornton Tomasetti's CORE studio assisted 360 architects in the panelization of the Rogers Place Arena in Edmonton, Canada. A bottom-up approach was used to derive panel layout controlled by physics engine Kangaroo for Grasshopper. © *Thornton Tomasetti CORE studio*

Excel of course is just an everyday tool. Everybody can program with Excel. But it is limiting, when looking at the larger picture, where we want all project information to feed into a central repository. Take for example a big stadium project. Recently there was a deadline. Two people from our team were involved. They spent four

Is how a tool handles data ever a factor in you considering working with these tools?

JS: Certainly. Let's start at the other end. Look at 3Ds Max. It's basically rendering software. Even if you were to measure the areas of its meshes, they wouldn't be accurate. Certain software is incapable of, and not meant for, data extraction and data processing.

Once Grasshopper came out, we found that it was good for nearly everybody—especially engineers, who are good at thinking logically; they write Excel functions and macros every day. Grasshopper is like Excel coupled with AutoCAD. They knew AutoCAD, they knew Excel. So this was just another way to combine data with geometry. I would say that Grasshopper is our #1 tool right now. It's so easy to say, "show me all the beams in the building and give me the ones that are longer than 5 feet." It is so easy to do that kind of analysis. [See Figure 1.33.]

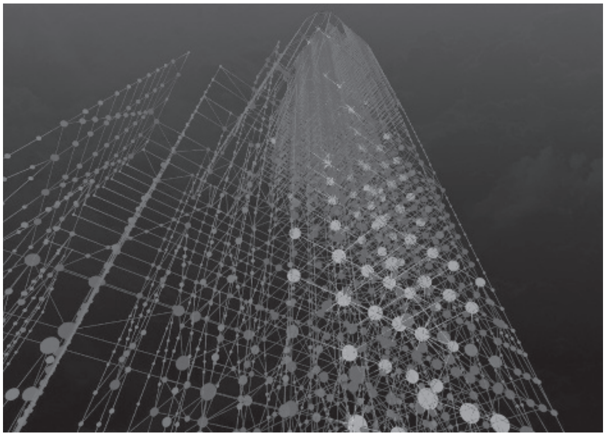


Figure 1.33: Thornton Tomasetti's in-house structural design suite: Thornton Tomasetti's CORE studio developed a number of tools for analysis of complex structures, and data visualization and mining thereof. © *Thornton Tomasetti CORE studio*

days working until 5:00 am in the office. Why couldn't we work smarter? They were like, this stair is being designed in that spreadsheet, and this part of the building is being designed over there. In the end, it's hard to combine everything into a single model. Everybody does their own separate thing. Nobody talks to the big repository of information. If used systematically, Grasshopper allows us to combine and mine information and data coming from different sources, such as spreadsheets and various BIM and analysis environments. But Grasshopper is still not a good database storage solution itself. This is why we developed TTX since Fall 2012. [See Figure 1.34.]

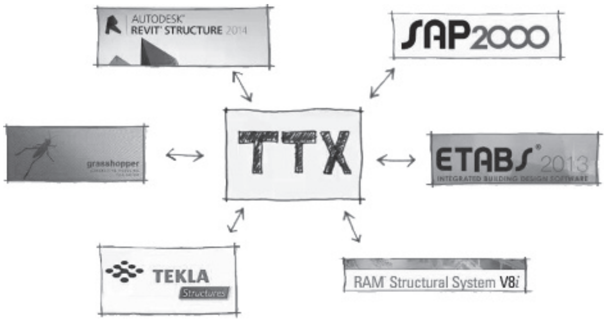


Figure 1.34: Thornton Tomasetti's CORE studio developed an in-house interoperability platform and BIM management suite: TTX. © *Thornton Tomasetti CORE studio*

Before deciding to develop our own interoperability platform, TTX, we were testing IFC file format on a large, fast-paced project. Certain companies, like Autodesk, are not motivated to work with IFC. We needed to get all this data from both Grasshopper and SAP into Revit, and it was not possible to do so in the workflow that the project required. If the input geometry changes, you lose track of which beams [in Revit] to replace with which beams [coming from Grasshopper]. IFC does not keep track of the unique identifiers that each program assigns to their BIM elements, so we can't use it well to make updates to existing models—especially if that model has changed, too. That is why we came up with TTX. It's an alternative to IFC. It's a file in the end, a database that contains all of the BIM information. It grows over time, and it can talk to all the different programs that we commonly use to model, analyze, document, and fabricate building structures. TTX is the common repository. We can now talk between the individual elements in all programs and keep updating our calculations. Over time we naturally keep growing this repository, as the project evolves.

In terms of finding talent, why would someone with a computer science background go to work in the AEC industry?

JS: Especially when it pays a third of what they were making in their respective industry… This person wanted to do some real, physical projects. We were lucky. There is obviously a large difference between creating software, or crunching numbers, and designing buildings that will live on for decades, which is attractive to some.

Do you see a need for exploring algorithms to further our capabilities and performance in design and construction over and above their capacity for generating form?

JS: In our R&D group, most people have a background in product design, engineering, or architecture, with a very strong interest in computer science. Very practically based. A couple of our people came from firms where they were working in Digital Project, or from a construction management or fabrication background. Computer science is important, as is an interest and expertise in a field related to our industry. It makes it hard for an engineering firm if the person only knows how to model well in BIM. That's not enough.

In terms of form-driven versus data-driven—both are nice challenges. On many interesting projects the architect doesn't necessarily think about data first and foremost. They're inspired by something formal. The data still represents a nice challenge and can be applied to any kind of design.

(*Continued*)

Some architects—big-name architects—don't care at all about the data. It is surprising. There are firms that tell us, we don't do 3D. It doesn't matter. Even then, as engineers we will do it for our own sake. We have to realize the geometry just as any engineer would. We just use different methods to get there.

We're very fortunate that our CEO Tom Scarangello made this conscious decision for Thornton Tomasetti to be the forerunner amongst engineering firms on the technology side. That's why we are investing heavily in R&D in our field. Tom understands that it will ultimately help the owner. We are often hired directly by owners, as opposed to architects or contractors.

We're focused on how buildings get built and what the complications will be on the construction side. This is why we want to run these kinds of studies during the design phase. Because there's a much greater likelihood that the building will get realized, compared to other high-end engineering firms that mainly work in the conceptual phases of a project.

Any other ways you can compress the process by using data or without losing the value of the data you already have?

JS: As engineers, long before all of this data talk, even before BIM was called BIM, we had 3D models with attributes. Data is always informing our designs. It is hard to address because I don't think of these as two different things. There is always geometry and data. The data is as important as the geometry is. The Petronas Towers in Kuala Lumpur, Malaysia, in the early nineties were analyzed in 3D. That's a BIM model. It's just that nobody had the term for it at the time. Data has always been a big part of what [structural] engineers do. [See Figures 1.35 and 1.36.]

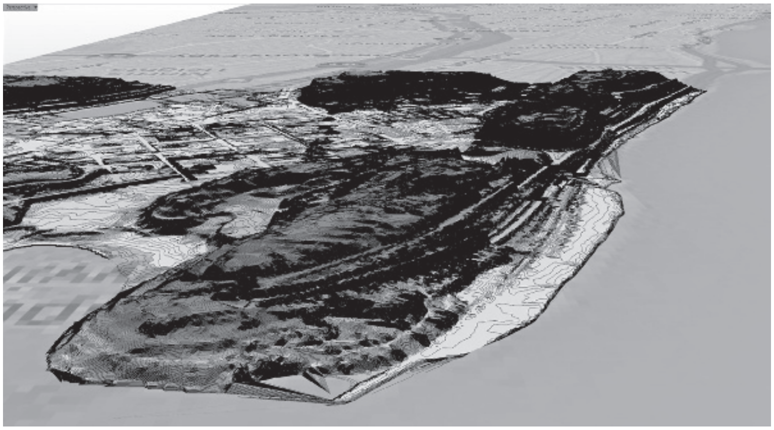


Figure 1.35: Hurricane Sandy disaster visualization: CORE studio assisted the Property Loss Consulting Group at Thornton Tomasetti in visualizing data captured after investigations. © *Thornton Tomasetti CORE studio*

You work closely to integrate the building structure, building skin, and building performance. How and where does data come into play?

JS: On a current project, the Hudson Yards Culture Shed by Diller Scofidio + Renfro, we are the structural engineers as well as the façade engineers. It's a kinetic structure where the structure and the skin are one and the same. The skin sits inside of the structural frame. If you were to try to coordinate between two different firms, it would be a nightmare to manage. We're also helping to algorithmically design the frit pattern that is printed onto the ETFE skin panels, and as such integrating sustainability services into the design process.

The structural model and the skin model on this project are the same thing. It is a geometrically complex kinetic structure, which will sit on top of the Hudson Rail Yards. It is important to coordinate the information so that all the disciplines can work with them. We're designing the frit patterns, for example, not just as an image, but with a set goal for reduction of a predescribed amount of solar radiation. This is something that our skin group would not be able to

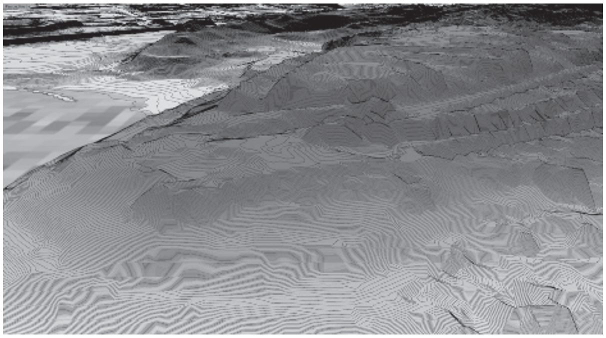


Figure 1.36: Hurricane Sandy disaster visualization: CORE studio assisted the Property Loss Consulting Group at Thornton Tomasetti in visualizing data captured after investigations. © *Thornton Tomasetti CORE studio*

do. Because they don't have the computational power to model and mesh all of these details. Our sustainability group wouldn't be able to do this either by itself. We have to integrate the knowledge of the different disciplines—the knowledge of materials, and of solar performance—and automate the creation of frit pattern, as well as the radiation analysis—it is a very computationally heavy process. There are a lot of analyses being run just to figure out what kind of frit pattern to use. We are doing this with Grasshopper—so we can make real-time adjustments as we go, and as the building geometry evolves—using Ladybug and Honeybee, two Grasshopper plug-ins that Mostapha developed. It is all parametrically linked.

The real-time data, in these instances, helps you to make more assured decisions. How do you communicate the data that supported your decisions to the architect/client?

JS: I hope that we will soon be able to communicate issues and design recommendations to our client—the owner or architect—in real time. See that red area there in the model? That we still need to fix. So let's just fix it now, in the shared model, in the web browser.

In the past, you'd have to throw out your previous iterations.

JS: Exactly. Now we can work with the same model. Now we have a parametric model, so we can change the geometry and retrigger analysis to be run. Our motivation has been to find ways to help the architects early on, really early on, in the process. So they can understand their building: How much does it cost? How much does it weigh? How will it be fabricated? These we answer in the structural analysis program. Now, with these visualization methods, we can comfortably go to the owner, convey our findings, creating trust from the beginning.

Another concept we are actively developing is what we call remote solving. This started in a conversation with LMN Architects tech studio (LMNts.) Traditionally, there is a huge disconnect between engineers and architects—especially during the early design phases. Engineers tend to wait for architects to "freeze" their designs, before they will even take a look—and then they will just post-rationalize it. The motivation behind remote solving is to be able to proactively inform the architecture, while it's being designed, with engineering and constructibility constraints. [See Figures 1.37 and 1.38.]

Currently, there is no ideal workflow defined for file exchange between A and E. So often, we are given a surface model, and we have to spend significant time to find a way to extract the centerline geometry from that. By the time we give them back the results the design has changed, and we are not able to inform the design in the early phases. So we came up with this: We are hosting the analysis model on a server, and expose certain inputs and outputs to the architects (and to other collaborators). Then, every time that the architect makes a change, the analysis automatically runs and provides feedback necessary for the architect to make an informed decision for their next design iteration.

(*Continued*)

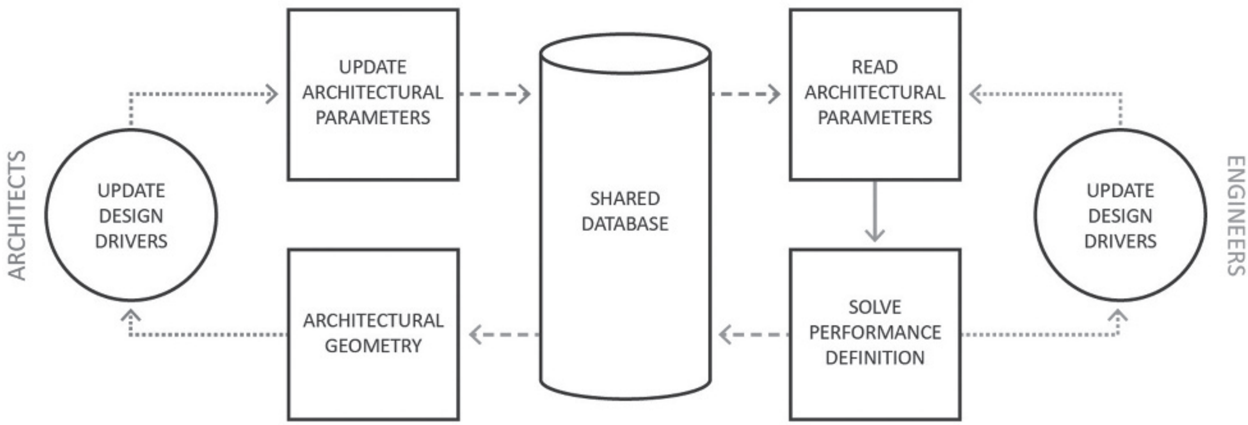


Figure 1.37: Thornton Tomasetti joint research project with LMN Tech Studio. Remote Solving allows for automated analysis feedback by engineers at concept design phase. © *Thornton Tomasetti CORE studio*

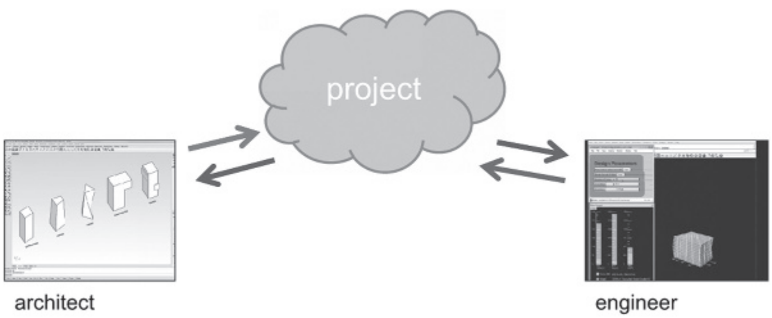


Figure 1.38: Thornton Tomasetti joint research project with LMN Tech Studio. Remote Solving allows for automated analysis feedback by engineers at concept design phase. © *Thornton Tomasetti CORE studio*

In this example, the architect can control the massing and the grid lines. Every time they make a change, the architect's computer uploads the new geometry to the database on Amazon's cloud. Our computer downloads that, resizes everything in real time, and a minute or two later they have their updated tonnages, structural sizes, and carbon values.

There are some firms that want to hire employees with data visualization skills. At TT, this would be superfluous. Your data viz is built into your system.

JS: Here's an example where these are the drivers and the architects could drive them themselves. The architects were interested in panelizing a double-curved façade surface in a way where every panel would have the same exact geometry. We developed a script to help them do this. Moreover, the façade engineer advised that we should check for curvature of the panels, and make sure that they don't warp more than 20 mm. So, as part of our script we measured deflection in real time, and visualized it in color (red = too much warpage). In doing so, we gave that script back to the architects, so that they could investigate different design options. They could drive how long they wanted the façade edges to be and what angle they wanted them to be. Based on their drivers, the façade would essentially push and pull itself into place. The goal again is to have as little red as possible. This way, they can see which angles work and which don't work. We embedded fabrication intelligence into their design model. That way they have the data and can figure it out themselves. You can go with any design you like. But if there's too much red, for example, it's going to be very expensive. [See Figure 1.39.]

Does TT collect and warehouse its own data for use in projects or to improve performance?

JS: As part of our intranet solution, we have a private webpage for every project that features high-level project information: who is the key contact, services offered, construction date, etc. We can use this intranet to ask: what do we do in healthcare, what do we do on high-rise projects, what do we do in Dubai? Every project page also has inputs for structural system, average building weight per square foot, and for embodied carbon. I have been considering adding the TTX model for every project in there, too. So that in the future, we can always look back and extract BIM and analytical data. It's just a database, so we'll be able to open and read it. It won't get outdated, like a Revit model or a Grasshopper definition would. And it doesn't use up much storage capacity. We can open it in 10 years and run very detailed queries down to a single BIM element or structural analysis node.

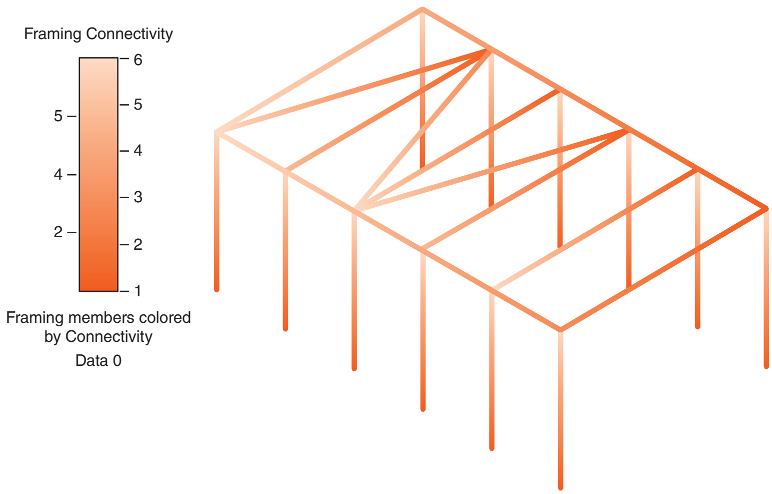


Figure 1.39: Thornton Tomasetti in-house structural design suite: Thornton Tomasetti's CORE studio developed a number of tools for analysis of complex structures, and data visualization and mining thereof. © *Thornton Tomasetti CORE studio*

There is a need today for a thorough understanding of how data is being leveraged in architecture, engineering, and construction, and by owners and operators. The innovative use of data in design and construction has been enabled by recent advances in technology and workflows, but also by access to information and an improved appreciation of how data can positively inform a variety of processes in the profession and industry.

Notes

1. Unless otherwise indicated, quoted text throughout the book is from interviews with the author that took place between February and July 2014. http://techcrunch.com
2. David Fano, interview with author, March 10, 2014.
3. David Fano, "BIM in the cloud: Industry view," *AEC Magazine*, July 29, 2014; http://aecmag.com/59-features/627-bim-in-the-cloud
4. Mads Jensen, interview with author, May 13, 2014.
5. Quote from Marina Gorbis, Executive Director, Institute for the Future, at an NBBJ-hosted event called "Data and Delight," Bloomberg San Francisco, August 18, 2013.
6. Randy Deutsch, quoted in David Barista, "The Big Data revolution: How data-driven design is transforming project planning," *Building Design and Construction*, February 11, 2014; http://www.bdcnetwork.com/big-data-revolution-how-data-driven-design-transforming-project-planning
7. Ibid.
8. Sean D. Burke, "Autodesk University, a ReCap," *Paradigm Shift*, December 18, 2013; www.seandburke.com/blog/2013/12/18/