TDWI

MONOGRAPH SERIES

MAY 2008

Second-Generation Collaborative Data Integration: Sustainable, Operational, and Governable

By Philip Russom
Senior Manager, TDWI Research
The Data Warehousing Institute

SPONSORED BY

sas

tdwi
THE DATA WAREHOUSING INSTITUTE

Table of Contents

About the Author

PHILIP RUSSOM is the senior manager of TDWI Research at The Data Warehousing Institute (TDWI), where he oversees many of TDWI's research-oriented publications, services, awards, and events. Before joining TDWI in 2005, Russom was an industry analyst covering BI at Forrester Research, Giga Information Group, and Hurwitz Group. He's also run his own business as an independent industry analyst and BI consultant, and was contributing editor with *Intelligent Enterprise* and *DM Review* magazines. Before that, Russom worked in technical and marketing positions for various database vendors. You can reach him at prussom@tdwi.org.

About Our Sponsor

Collaborative data integration is a key component of the business intelligence solutions offered by SAS, the leader in business intelligence and analytical software and services. Customers at 42,000 sites use SAS software to improve performance through insight from data, resulting in faster, more accurate business decisions; more profitable relationships with customers and suppliers; compliance with governmental regulations; research breakthroughs; and better products and processes. Only SAS offers leading data integration, storage, analytics, and business intelligence applications within a comprehensive enterprise intelligence platform. Since 1976, SAS has been giving customers around the world THE POWER TO KNOW®. www.sas.com

Defining the Second Generation of Collaborative Data Integration

TDWI Research defines collaborative data integration as:

> A collection of user best practices, software tool functions, and cross-functional project workflows that foster collaboration among the growing number of technical and business people involved in data integration projects and initiatives.

The first generation of collaborative data integration came into existence for pragmatic reasons: [1]

The first generation reacted to increasing DI team size, geographic dispersion, and links to related practices.

Data integration specialists are growing in number. Collaboration requirements intensify as the number of data integration specialists increases. Many organizations have moved from one or two data integration specialists on a data warehouse team a few years ago to five or more today.

Data integration work is increasingly dispersed geographically. Projects that involve data integration are progressively outsourced. Even when the entire project team works for the same organization, employees may work from home, from various offices, or while traveling.

Data integration is now better coordinated with other data management disciplines. Data integration specialists must coordinate efforts with specialists for data quality, metadata management, data warehousing, master data management, operational applications, database administration, and so on. All these specialists experience moments where they must work together or simply have a read-only view of data integration project artifacts.

The second generation of collaborative data integration addresses a relatively new set of trends:

The second generation is driven up by DI's growing role in sustainability, system consolidations, and governance.

The system consolidations typical of a green data center require data integration. Recent climate changes and the rising price of electricity have led many people to revisit the sustainability of data centers. In response, many corporations are reducing power consumption and the physical footprint of data centers by consolidating and virtualizing redundant data and hardware servers. Most consolidations require that data be transformed and cleansed to better fit the target, which demands tools and techniques for data integration. So, data integration teams are collaborating with green data center programs to effect the consolidations that make IT more sustainable.

Operational data integration is inherently collaborative, and it's growing as a practice. Operational data integration aggressively consolidates, migrates, collocates, and upgrades systems, usually in the context of a green data center program, ERP initiative, IT centralization, or merger. You can't kill and replace systems successfully without detailed collaboration with the business units that use those systems and the business managers who own and fund them.

Data integration's collaborations are increasingly structured by data governance. Business people now need direct oversight into data integration, due to new requirements for compliance and governance. And data integration is one of the first data management practices that a new data governance board controls, because data integration work effects changes that reach across multiple business units and programs. For these and other reasons, the collaborative side of data integration is progressively enabled and controlled by cross-functional data governance boards.

This TDWI Monograph looks into the new set of trends that are driving the second generation of collaborative data integration.

---

[1] For a discussion of the first generation, see the TDWI Monograph *Collaborative Data Integration: Coordinating Efforts within Teams and Beyond*, available for download on the Monograph Series page under Research at www.tdwi.org.

Collaborative Data Integration and Sustainability

Recent changes in our planet's climate and the rising cost of energy (which is generated from the Earth's resources) have gotten a lot of attention in recent years. These dramatic events have prompted many people worldwide to question whether human civilization will be sustainable in the long term given its negative impact on the planet's ecology and resources. Discourse on these issues has lead to the concept of sustainability, defined as follows:

> Sustainability is a characteristic of a process or state that can be maintained at a certain level indefinitely. The term, in its environmental usage, refers to the potential longevity of vital human ecological support systems, such as the planet's climatic system, systems of agriculture, industry, forestry, and fisheries, and human communities in general and the various systems on which they depend in balance with the impacts of our unsustainable or sustainable design.[2]

Sustainability involves both ecological and economic systems.

In this definition, sustainability was conceived with the planet's natural environment in mind, which is why sustainability is regularly associated with "green" concepts. Yet, sustainability also applies by extension to other, smaller environments that aren't necessarily green, namely the homes, farms, businesses, and other societal units of human civilization. For example, new practices in sustainable home construction use recycled or recyclable materials, and sustainable agriculture ensures that land will be farmable for many generations. Note that these practices help sustain the ecology, while also sustaining the economic dynamics of civilization.

Relevant to this TDWI Monograph, data centers and related IT infrastructure consume an increasing amount of resources annually, especially electricity. A recent Energy Star study estimates that U.S. data centers "consumed about 61 billion kilowatt-hours (kWh) in 2006 (1.5 percent of total U.S. electricity consumption) for a total electricity cost of about $4.5 billion … The energy use of the nation's servers and data centers in 2006 is estimated to be more than double the electricity that was consumed for this purpose in 2000 … Under current efficiency trends, national energy consumption by servers and data centers could nearly double again in another five years (i.e., by 2011) to more than 100 billion kWh."[3]

With data center electricity consumption doubling every five years, it's no surprise that many businesses, government agencies, and educational institutions are starting to apply sustainability concepts to their data centers—for both ecological and economic reasons. Today, IT personnel mostly measure the sustainability and ecological impact of the facility that houses the data center and its hardware infrastructure. Less often, IT personnel measure individual software applications and the data managed by them—areas that are in dire need of sustainability improvement.

Data management is one of the least sustainable areas in IT.

In particular, the broad field of data management is suffering onerous trends that challenge the sustainability of data center resources:

- **Exploding data volumes.** As enterprises continue to automate as many processes as possible with software, numerous online systems collect data at unprecedented and increasing rates.
- **Unrealistic retention policies.** Driven by regulatory and legal requirements (and audit paranoia, too!), organizations retain almost all data, regardless of its age or usefulness.

---

[2] Source: Wikipedia at http://en.wikipedia.org/wiki/Sustainability.

[3] Source: *Report to Congress on Server and Data Center Energy Efficiency* by the U.S. Environmental Protection Agency's ENERGY STAR Program, published August 2, 2007.

- **Rising costs.** Massive data volumes retained indefinitely demand server and storage hardware, administrative personnel, data center space, network capacity, and electricity.
- **Ecological impact.** Resources gobbled up by data management increase the energy input and carbon output of the data center.

Green data centers use data integration extensively.

A green data center is a data center that seeks to be as sustainable as possible. Similar terms include green IT (which doesn't necessarily involve a recognizable data center) and green computing (which tends to focus on the greenness of hardware manufacturing and recycling). Synonyms arise when the term "sustainable" replaces "green."

Green practices for the data center vary tremendously. At the low end, some organizations found a green data center program that simply tidies up the data center by consolidating a few redundant servers. At the high end, many major corporations have recently broken ground to construct new data center facilities that are green from the ground up and designed for maximum sustainability. Most programs fall between these two extremes, where an existing data center is treated to numerous system consolidations and retrofitted with ecologically friendly computer hardware and cooling systems. The consolidations and retrofits typical of a green data center seek to reduce both environmental impact and business costs.

Server consolidation is green, and it requires data integration.

System consolidation is one of the cornerstones of a green data center. Almost all green data center programs focus on the consolidation of applications and databases. The assumption is that various types of system consolidations (and related projects, like system migrations, collocations, and upgrades) can reduce the number of hardware servers and storage devices, which in turn reduces power consumption and the data center footprint. Consolidations may also have non-green drivers, like tool standardization (where brands that are not the corporate standard are eliminated), instance reduction (where multiple instances of the same brand are merged into one), and centralized administration.

Virtualization is a viable alternative to the consolidation of some servers. Computer virtualization involves running two or more logical software applications on one set of physical hardware. Virtualization goes back to mainframes, but was resurrected in the 1990s and adapted to open system servers to get greater efficiency from collocated software servers. Whereas consolidation merges homogeneous software servers onto one hardware server, virtualization can collocate heterogeneous software servers. Hence, the two are complementary and therefore both are useful weapons for the green data center's arsenal.

A short-term investment in new power-efficient hardware may be in order for long-term benefits.

Power-efficient hardware is the other cornerstone of the green data center. Enterprise-class hardware that is power-efficient is relatively new, which means that the vast majority of server, storage, and network infrastructure in the average data center today is not power-efficient. And there don't seem to be solutions that retrofit energy management onto older systems. Hence, achieving a significant reduction in power consumption may mean swapping old hardware for new. Hardware swaps often coincide with data integration work that transforms data to enhance its performance on the new platform or to improve data's model and quality.

Real estate also impacts cost and ecology.

A green data center also reduces its real estate footprint. Extra square footage occupied by the data center means extra electricity, air conditioning, lighting, construction, maintenance, and so on. All these have impact on both business and ecological sustainability. Data integration's consolidation projects contribute to footprint reduction, so the two should be coordinated.

Data integration helps data centers go green.

The system consolidations typical of a green data center require data integration. And that's also true of similar project types, like the migration, collocation, or upgrade of applications and databases. Although a few application and database consolidations may be executed by simply copying data from source A to target B, the vast majority require that data be transformed and cleansed to better fit the target. And that demands tools and techniques for data integration.

Data integration is preferred over virtualization, at least in the data tier. Virtualization is most often applied to application servers, which can be collocated and tweaked in a straightforward manner to share a common memory space and other hardware resources. Data servers are a different matter entirely, because all are designed to seize every scrap of hardware resource. Let's recall that a data server (which is a software server) differs from a database (which is where the data resides); the former manages the latter. While users regularly collate multiple databases successfully, the collocation—much less virtualization—of multiple data servers (call them database management systems, if you prefer) is unlikely to yield desirable results. Therefore, in the data tier of the green data center, database consolidations and collocations—which are best done with data integration techniques—are preferred over true virtualization.

A green data center reduces both business cost and ecological impact. Achieving either goal is unquestionably worth doing. Achieving both in one fell swoop is one of the most compelling opportunities ever presented to an enterprise. And data integration is a primary technology used in achieving these worthwhile reductions.

Operational Data Integration is Sustainable and Collaborative

Let's pause for a moment and define two broad categories of data integration, which are distinguished by the types of projects with which they are associated:

- **Analytic data integration** is applied most often to data warehousing (DW) and business intelligence (BI), less often to initiatives that resemble these, like customer data integration (CDI) or master data management (MDM).
- **Operational data integration** manifests itself as projects or implementations commonly described as the consolidation, collocation, migration, upgrade, or synchronization of operational databases.

As you can see, operational data integration executes the types of application and database consolidations commonly found in programs for green data centers. Hence, operational data integration is a significant contributor to sustainability, so these programs must collaborate with data integration specialists.

SURVEY SAYS:
At 25–39% of workload, operational DI is an established practice.

Three times, TDWI Research has surveyed TDWI Members, asking: "With ETL usage in your organization, what is the approximate split between ETL applied to data warehousing and ETL applied to other uses (like database consolidations or migrations)?" Responses have varied, revealing between 25% and 39% of ETL usage devoted to database consolidations and similar projects. This proves that operational data integration is an established practice that's a sizable chunk of many data integration specialists' workload.

Operational data integration addresses sustainability issues.

Operational DI targets redundant or non-standard applications and data, which are not sustainable.

Redundant data and nonstandard databases are the main problem. For example, when data repeats across multiple databases, it's hard to keep the databases synchronized. Likewise, data may reside in a legacy database that is beyond its prime or in a database brand that is simply not the corporate standard. To put it another way, these are problems because they increase IT costs and inhibit unified visibility into business processes. In a related trend, many organizations centralize both IT and business operations. These situations eschew redundancy and promote standards; thus they seek solutions via database consolidations, migrations, and so on.

Redundant and nonstandard applications are a problem, too. Consider an operational application that's a legacy needing migration to a more modern brand or an application instance that should be consolidated with other instances. When there's a database in the application's technology stack, operational data integration is required to migrate or consolidate the database.

Redundant or nonstandard data and applications are relatively nonsustainable. As discussed earlier, green programs actively target redundancy under the assumption that redundant systems increase resource consumption and carbon load.

SURVEY SAYS: ETL is the preferred technology for operational DI.

Operational data integration is most often executed with ETL. Twice in 2007, TDWI Research surveyed TDWI Members, asking: "When your organization consolidates or migrates databases, what is the preferred technology for most projects?" Both times, roughly half of survey respondents selected extract, transform, and load (ETL), far more than selected hand coding, replication, and other techniques.

Operational data integration is inherently collaborative.

Operational data integration always involves some kind of collaboration. This is true whether operational data integration is supporting a green data center program, data warehousing, IT centralization initiatives, ERP, or other application initiatives.

Data integration specialists rarely choose which systems get consolidated. These decisions are invariably made through a collaboration of many people, including the data integration specialists, people from related teams (like data warehousing or database administration), line of business managers, the CIO's office, and others.

Operational data integration takes away systems and introduces new ones. You can't do that without careful collaboration with the business units that use those systems and the business managers who own and fund them. The collaboration is a mix of gathering business requirements up front, as well as coordinating application and database switch-overs and hand-offs later.

Operational data integration is rarely a quick, one-shot deal. These projects usually entail multiple phases and steps (as the user stories below will illustrate). And the "old" system that's being phased out may continue running for years, until all its functionality is migrated, tested, and approved by users and owners. In fact, it's normal that old and new systems run simultaneously for some time, with data integration between them to keep them synchronized. Planning, executing, and coordinating all these steps demands much collaboration.

Operational data integration should improve data, not merely copy it to a new platform. It should improve the quality of both physical and semantic data (like metadata and master data). In many projects, the data model benefits from updating, although change may not be possible when the target must comply with a prescribed data model (as is the case with most packaged applications). The more "legacy" the source data is, the more needed are improvements. Identifying, specifying, and making improvements require collaboration between data integration specialists and others.

USER STORY: Multi-year, multi-step operational DI projects require intense cross-functional collaboration.

Global manufacturer. Early this decade, a global manufacturer decided to consolidate its 17 ERP instances into three, as a way of gaining tighter business integration and reducing costs. The catch was that direct consolidation was not possible, because the ERP instances were of diverse vendor brands and release versions, running on diverse hardware and operating system platforms. Furthermore, the instances were physically located in offices around the globe and owned by regional managing directors. Before consolidation could occur, the instances all needed to be migrated to the same brand and upgraded to compatible versions. Collocation was a practical interim step along the way.

It took over a year of intense collaboration among technical and business people to devise a multi-step plan for all the steps involved in operational data integration and equivalent tasks for application consolidation. Besides technology steps, the plan also established a schedule for switching application end-users to various upgraded and migrated application instances, since the company needed uninterrupted application use to keep business operations running.

After three years of execution, the manufacturer achieved its goal of three ERP instances, one for each of its major geographic regions. The reduction of instances helped them save millions in IT costs, see a fuller picture of the international business, gain greater business agility, and become more sustainable from both business and ecological viewpoints. Figure 1 approximates the complex order of operational data integration projects involved in consolidating 17 ERP instances into three. This user story illustrates that highly complex problems require complex, multi-step projects with multiple technical solutions and a lot of cross-functional collaboration.

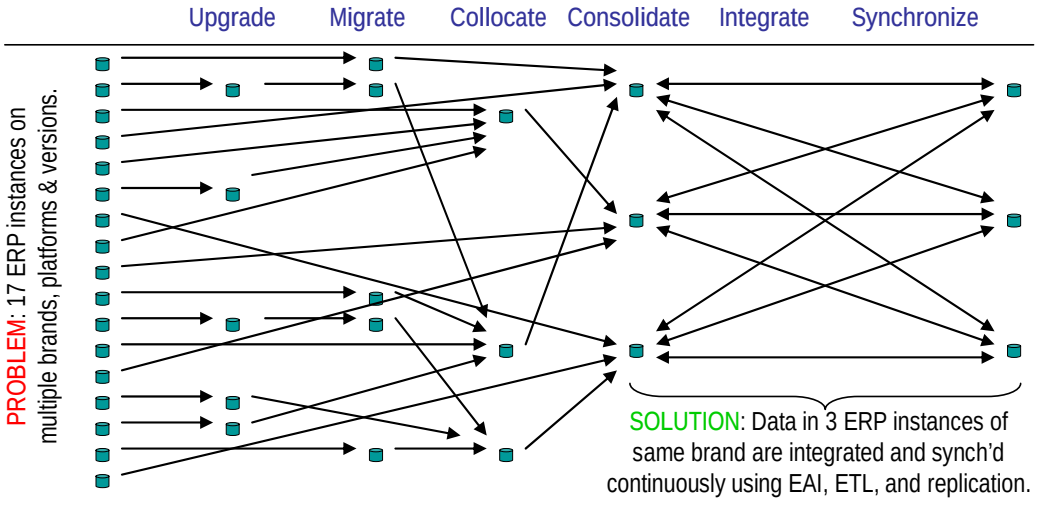

Figure 1. Source: TDWI Research.

USER STORY: Recurring M&As mean recurring operational DI projects.

Regional Retail Bank. The U.S. financial industry has seen thousands of mergers and acquisitions since it was deregulated in the 1990s. In particular, retail banks have acquired other, smaller banks in an effort to grab greater market share and to grow via acquisition. Indications are that the flurry of mergers and acquisitions (M&As) will continue.

Recurring M&As are challenging for data integration teams. Consider a regional chain of retail banks located in the U.S. For years, this bank has been making an acquisition every 12–18 months or so. The operational data integration tasks are handled by the data integration specialists on the data warehouse team. The data integration specialists do a great job because of their technical skills honed in data warehousing. They also have excellent cross-functional collaboration skills, again learned from years of data warehousing, which is inherently collaborative. Hence, the data integration specialists have no trouble handling the data or working with business people to coordinate the hand-offs in application and data use and ownership that are typical of a merger. And the data integration guys say that it's best they're involved, so they have control over the quality of data that will also feed the data warehouse. Figure 2 approximates the complex operational data integration steps involved in a typical merger or acquisition.

The problem, however, is that data integration specialists must put their data warehouse work on the back burner, while they deal with yet another M&A. Even worse, the M&As come unpredictably, and each one is different. After years of M&A activity, the work of the data warehouse team is severely behind schedule, and the whole company is suffering from inadequate business intelligence.

But there's light at the end of the tunnel. Since the bank's management anticipates many more acquisitions, it recently approved a data integration competency center. Eventually, the competency center will be the home of an expanded data integration team that's independent of the data warehouse team. Once it's in place, the competency center will collaborate with the data warehouse team, M&A team, green data center program, IT in general, and other business units, following a shared services model.

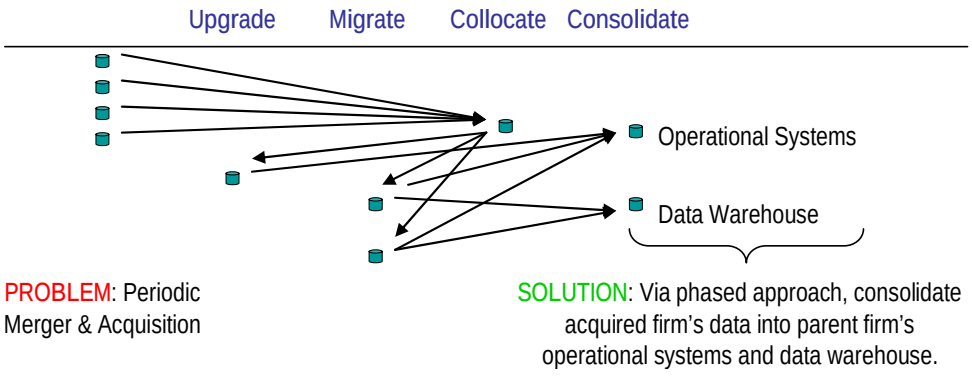

Figure 2. Source: TDWI Research.

Collaborative Data Integration and Governance

Collaboration requires structure to get everyone organized, and certain organizational units can provide that structure for data integration. These organizational structures can be technology focused (like data management groups), business driven (like data stewardship programs and governance boards), or a hybrid of the two (like BI teams and competency centers). More and more, however, the structure for collaborative data integration is some form of governance, including corporate governance, IT governance, data governance, BI governance, or a hierarchy of governance boards that federates all these.

SURVEY SAYS:
The leading benefits of collaborative DI are governance support and visibility into DI work for business people.

TDWI Research surveyed TDWI Members in May 2007, asking: "What are the benefits of collaborative data integration?" Survey respondents identified the leading benefit as "[collaborative data integration] supports data governance and stewardship requirements" (71%), followed by "business sponsors have visibility into data integration" (55%).

The two benefits are closely related. Data is so central to many compliance requirements that governing bodies need to literally look into data integration projects to understand whether they are compliant. For example, one of the typical responsibilities of a data governance committee is to ensure that data for regulatory reports is drawn from the best sources and documented with an auditable paper trail. Goals like this are achieved faster and more accurately when supported by the "big picture" provided by a combination of collaborative data integration and governance.

Second-generation collaborative DI reaches beyond BI and DW into large-scale business initiatives.

From these examples, we see that data integration is no longer limited to first-generation collaboration within burgeoning data integration teams or with related teams, like those for business intelligence, data warehousing, and data quality. Instead, the second generation of collaborative data integration also reaches "up org chart" teams, namely purely business-driven initiatives that depend on data integration for success, like business integrations (e.g., 360 degree of customer, data as an enterprise asset) and business transformations (M&As, reorgs). These initiatives are inherently collaborative on the business level, since they involve coordination among upper management and a diverse line of business managers. Yet, data integration is a critical success factor in the execution of these initiatives.

SURVEY SAYS:
DI is the technology most likely to be controlled by data governance.

Given the complexity of these organizational combinations—and the inevitable disconnects of their extreme cross-functional communication—data integration's collaboration with business-driven initiatives is best done through some form of governance board, usually data governance. In fact, data integration for any purpose is progressively governed. To quantify the situation, in August 2007, TDWI Research asked TDWI Members: "Which data management practices do you think should be guided by data governance?" At the top of the list, survey respondents ranked data integration (83%), followed closely by data quality (82%).[4]

Second-generation collaborative DI is often about effecting dramatic change.

The reason that second-generation collaborative data integration is progressively driven into the heart of grander business initiatives and governance is that all three share common attributes, goals, and barriers. And most of these boil down to *change*.

Business integration forces owners to open up their data for broader sharing. When departmentally owned systems are involved, their owners and sponsors may dig in their heels and resist. Instead of the data integration team butting heads with them, the business integration team should work with a governance board to authorize and enforce the change of data access policy.

[4] For a detailed study of data governance, see the 2008 TDWI Best Practices Report *Data Governance Strategies*, available online at www.tdwi.org/research/reportseries.

Most business transformations involve data transformations, too. For example, organizational chart changes are often supported by data consolidations and migrations. A common one is the change of ownership of data seen in reorganizations or initiatives for data as an enterprise asset. M&As change the ownership of entire companies, affecting everything all the way down to data ownership and how data of the acquired company should be integrated with that of the parent firm.

Quality improvements force owners to change their applications and how they're used. Data quality and master data management programs force changes in the quality of data and its definitions. Typically a technical team proposes changes to the managers who own the applications or data that need improvement. A change of this type—which is both cross-functional and cross-business-unit—is unlikely without a lot of collaboration and governance.[5]

Green IT is inherently transformational. In most organizations, achieving green IT goals involves numerous nontrivial changes to software applications and data, as well as the people who own, use, design, deploy, and maintain these. With so much change to manage and so many people's mindsets to realign, it's best to involve governance, especially corporate governance (from which most green mandates originate). Data governance is also relevant, given that many green changes focus on data storage and operational data integration.

Data governance *is* a change management mechanism. The board provides a cross-functional collaborative environment for defining and organizing the dramatic changes listed above. And the procedures established by the average data governance board provide a pragmatic mechanism for proposing, reviewing, and policing cross-business-unit changes to data, its quality, and its use.

Summary and Recommendations

Collaboration is a real requirement for DI, and it's driven up by DI's growing role in sustainability, system consolidations, and governance.

Recognize that collaborative data integration is entering its second generation. The first generation was driven into existence by a need to coordinate large numbers of data integration specialists, manage geographically dispersed data integration teams, and collaborate with teams in data management areas related to data integration. The second generation is now emerging, due to data integration's deepening role in the system consolidations of green programs, the continuing growth of operational data integration practices, the rising number of business integration and business transformation initiatives (which require data integration as a critical success factor), and new collaborative processes available via data governance.

Put your passion into green DI work, and collaborate to create greater sustainability.

Work toward sustainability and greenness with all you have, including data integration. According to the EPA's Energy Star Program, the electricity consumption of data centers and related IT infrastructure is doubling roughly every five years, which means spiraling costs for user organizations. Green programs for data center operations can significantly reduce power consumption, which in turn reduces both bottom-line costs and carbon loads in the environment. In fact, to make sustainability as palatable as possible to a business audience, it should always be proposed as a balance of economic and ecological concerns.

Reinvigorate your passion for data integration by championing sustainability. Rarely do people working in the depths of data management have such a stellar opportunity to effect positive change on a grand scale that reaches beyond the glass house of the data center to global issues like the environment. The consolidations and virtualizations that data integration does in support of a

[5] The intrinsic relationship between data quality and data governance is explained in the TDWI Best Practices Report *Taking Data Quality to the Enterprise through Data Governance*. Likewise, the relationship between MDM and data governance is explained in the TDWI Best Practices Report *Master Data Management: Consensus-Driven Data Definitions for Cross-Application Consistency*. Both reports are available online at www.tdwi.org/research/reportseries.

green data center program is noble work that should fire your passions. And it gives you ample cause to leave your cube and collaborate with colleagues to develop new and creative ways to make IT and the rest of your organization more sustainable.

Consider a renewed investment in resources for operational DI.

Expect to expand your operational data integration practice. According to data developed by TDWI Research, the two broad practices of data integration—analytic and operational—are both growing, but the operational practice has grown faster than the analytic one, in recent years. This trend will continue, driven by the increasing number of initiatives that require operational data integration, including green data center programs, ERP initiatives, application consolidation initiatives, IT centralization, legacy decommissions, data as an enterprise asset programs, corporate reorganizations, and mergers and acquisitions. Plan ahead by staffing data integration functions so they can fulfill growing workloads in both analytic and operational practices. Consider deploying a data integration competency center that objectively provides shared services for both practices from a single, central, collaborative organizational unit.

Data governance is in the future of all data management, including collaborative DI.

Embrace data governance for its collaborative and cross-functional benefits. A view that many users share is that it's nonproductive to institute cross-functional collaboration processes for data integration, data governance, and related collaborative practices, like business intelligence, data warehousing, and data quality. Instead of reinventing the wheel with potentially contradictory processes, it might be better to drive all cross-functional and cross-business-unit collaboration through data governance. If this viewpoint wins out, it's possible that data integration will remain collaborative, but much of the collaboration will be structured by the policies and procedures of a data governance board. This makes sense, because data governance boards wield great clout that can help further the causes of enterprise data management. TDWI's recommendation (based on the actions of our own Members) is to embrace data governance and leverage its clout and procedures to enhance the collaborative side of data integration and related data management practices.