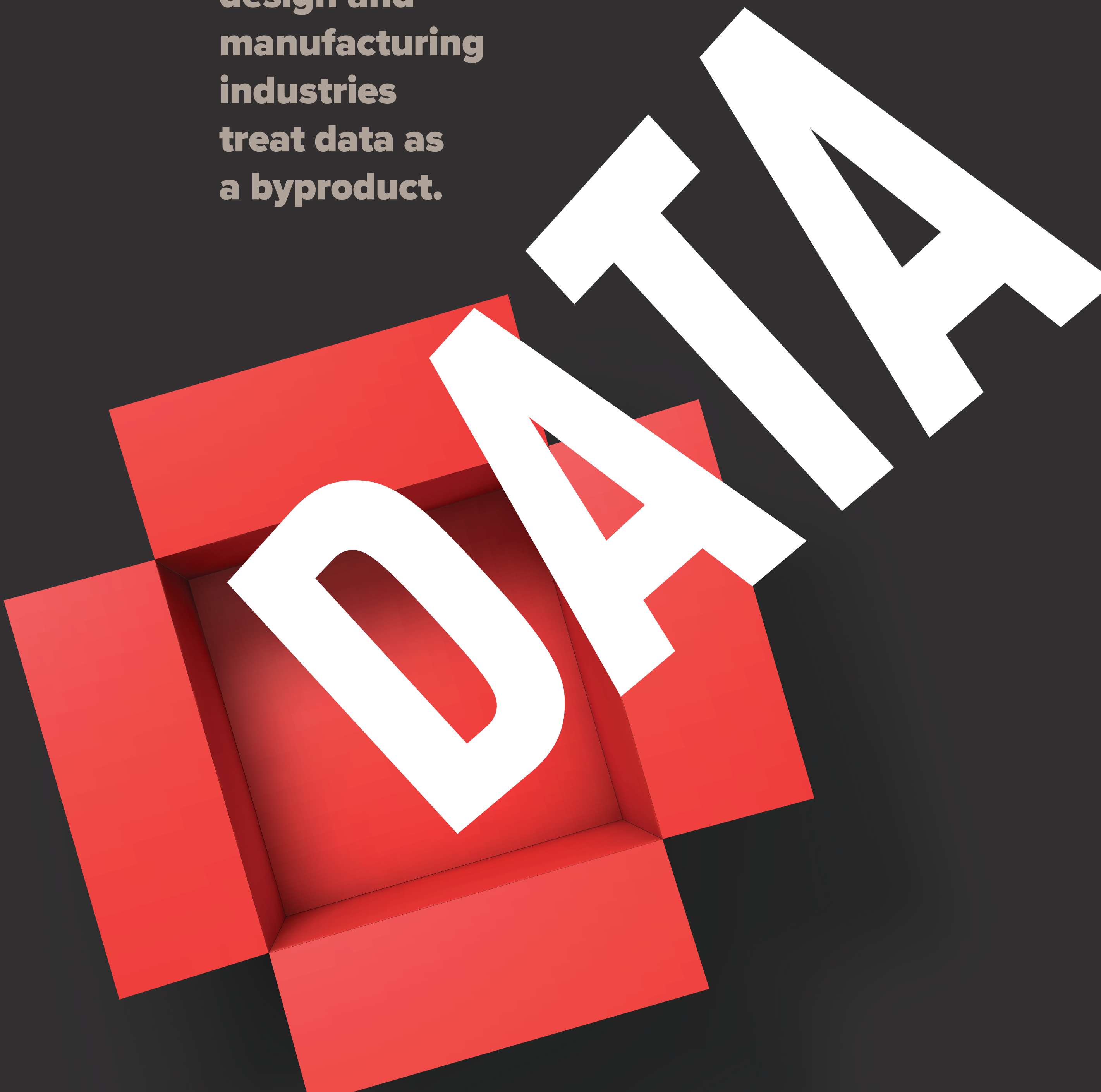
Today's design and manufacturing industries treat data as a byproduct.



Data comes out of storage.

AND MANUFACTURING INNOVATION

Many of today's designers and manufacturers view data that's generated during the development of a new product or manufacturing technique as a mere byproduct of those processes.

As a result, only the most rarified of the data produced during design and manufacturing processes is curated in digital formats that make it accessible and meaningful. Too often, we leave potentially valuable data in a state that realizes no current or future value. Enterprises of all sizes orphan important data on the shop floor.

This is a lost opportunity. A sufficiently rich data set that is fully accessible enables designers to discover previous processes and leads—including false starts and dead ends—that could develop into new solutions. Rather than throwing out this valuable data or leaving it in inaccessible forms, industry, researchers, and others may soon be able to use tools to explore this information and amplify their intelligence and experiences.

Before we can get to that point, though, we have to rethink the relationship between data and manufacturing innovation. We will have to understand that data is the central and most essential product of engineering design activity.

By William Regli

"DATA IS NOT A MERE PRODUCT OF PRODUCT LIFECYCLE ACTIVITIES—IT IS DATA THAT GIVES RISE TO THESE ACTIVITIES IN THE FIRST PLACE."

Keeping data isn't enough.

Data can be meticulously archived but also rendered utterly useless. For instance, it could be kept on paper or in an analog data format such as old Applicon files printed to aperture-style punch cards. Digital data stored in unsupported storage technology, such as tapes or floppies, is just as inaccessible. Digital data could be in a lossy or derivative format, such as a 3-D CAD drawing archived as a 2-D PDF, or it could lack the context or metadata to make it discoverable.

Another element usually missing from stored data is the thought process behind its creation. Design produces many branches that—as a collection—can be valuable; yet those design decisions, explorations, R&D tests, and alternative analyses are typically discarded.

While industry decision makers recognize that product and manufacturing data is important, they often lack an understanding of what constitutes product-related data and the actual value of that data.

For instance, industry today is rapidly adopting something called the

Industry today is rapidly adopting something called the DIGITAL THREAD, a best-practice in product data management that emphasizes a closed-loop process in which product lifecycle data is linked and traceable to design intent.

digital thread, a best-practice in product data management that emphasizes a closed-loop process in which product lifecycle data is linked and traceable to design intent. According to a definition published by the Defense Acquisition University at Fort Belvoir, Va., which provides acquisition, technology, and logistics training for military staff and contractors, the goal of the digital thread is to "inform decision makers throughout a system's lifecycle by providing the capability to access, integrate, and transform disparate data into actionable information."

The Digital Manufacturing and Design Information Institute (DMDII) has proposed that the digital thread serve as the backbone with which to aggregate and use manufacturing data. That would create a seamless flow of data across the product lifecycle and the information systems that manage and use this data. Some experts speculate that the thread will enable data analytics and reduce the time and cost to design and manufacture a physical product. Organizations such as the National Institute of Standards and Technology (NIST), GE, and many others have written about the promise of the thread to transform manufacturing into a fully digital enterprise.

In this way, the digital thread supports the requirements of the conventional manufacturing supply chain and the organizations that lead them. Yet in this Internet-of-things enabled supply chain, the data of interest is often limited to the very narrow digital thread that ties together the product data, production data and lifecycle information for a finished artifact.

These best practices in design and manufacturing stand in stark contrast to data revolutions being fomented in other fields. Beginning in the late 1990s, for instance, several large research initiatives have transformed science. In those cases scientists gather observations and share them using a data substrate, and researchers can verify analyses quickly and (most importantly) pose entirely new and previously unframable questions. The Sloan Digital Sky Survey, the Large Hadron Collider, the Ocean Observatories Initiative, and the iPlant Collaborative have helped turn astronomy, physics, oceanography, and biology into data-driven sciences. The Simplifying Complexity in Scientific Discovery program at the Defense Advanced Research Projects Agency (DARPA, the organization I work for) is also developing mathematical frameworks and tools for scientific data analysis in a variety of domains.

Recent accomplishments by a multidisciplinary team led by the American Museum of Natural History are changing the field of anthropology. The team "datafies" unstructured and semi-structured data from a disparate set of sources and in a diverse set of file formats into a unified knowledge representation schema. The hope is that this effort will enable scientists to pose questions about causation and correlation across multiple linguistic, genetic, and geographic data sources.

DARPA's Big Mechanism program employs machine reading technologies to enable computers to process and understand the results in tens of thousands published cancer biology articles. With this capability, scientists will be able to design new experiments of unprecedented acuity to unlock the larger machinery of the proteins that affect cancer growth. DARPA plans to roll out similar but much larger programs.

We now have the opportunity to reimagine design and manufacturing innovation as a data-driven exercise. The digital thread is simply the digitization of the existing process. To really create a revolution we must widen our aperture and rethink the basic nature of design and manufacturing activity in the light of the revolution in data and computation.

In this new thinking, the data ecosystem is the collective memory of the complete output of a project, from conception to post-production and lifecycle activity for its artifact. It must encompass the dead ends as well as the amazing insights; the unbuilt

The digital thread is simply the digitization of the existing process.

alternatives and the simulation tests for prototypes; the information arcs traced out in the course of design and manufacturing process; the consumer reviews and in-service information feeds from devices in the field.

Today's notion of the digital thread represents merely one strand in a vaster data substrate that is the landscape of the creative activity associated with design and manufacturing.

In this vision, data is the basis for all design, manufacturing, and life-cycle activities. New artifacts are produced out of information from—and interaction with—the data substrate. Production and manufacturing are the physical realizations of information on this substrate. Data is not a mere product of product of a life cycle—it is the information that gives rise to these activities in the first place.

Fundamentally, design and manufacturing are about the transformation of information into something tangible. Think of the process of publishing the magazine you are reading. Printing and binding the magazine is the final manifestation of the writing process that created the articles, the marriage of text and images via production software and the creative and business decisions made by dozens of professionals. The physical artifact of the magazine is all of that information.

This will become increasingly true of every manufactured object. 3-D printers already enable the rendering of objects not just as physical artifacts but as computer files containing all of the information necessary for their production. As advanced manufacturing becomes more prevalent, products will be seen as the physical embodiment of a vast sea of data containing

In this new thinking, the data ecosystem is the collective memory of

The **COMPLETE OUTPUT OF A PROJECT,** from conception to post-production and lifecycle activity for its artifact

The **DEAD ENDS** as well as the amazing **INSIGHTS**

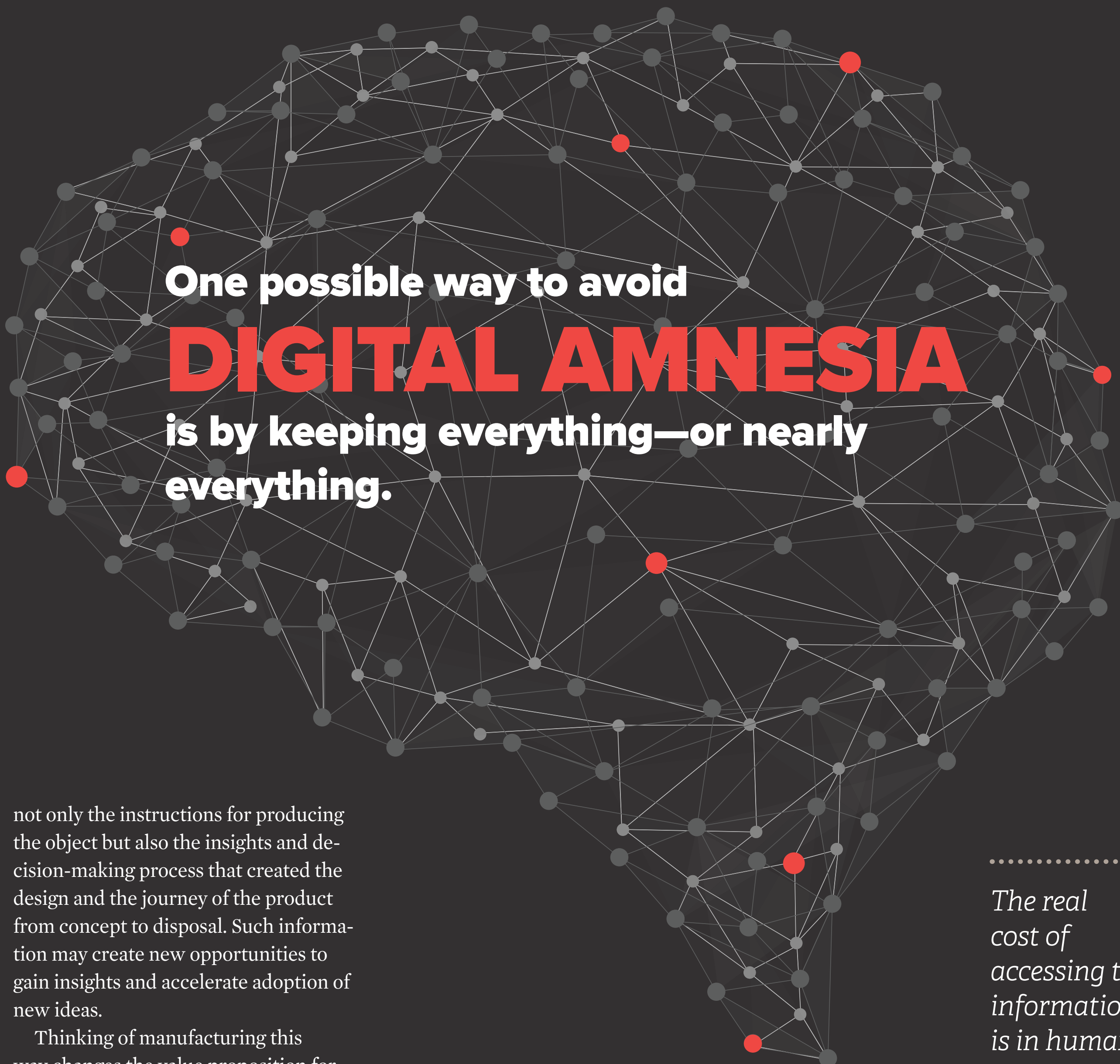
The **UNBUILT ALTERNATIVES** and the **SIMULATION TESTS** for prototypes

The **INFORMATION ARCS** traced out in the course of design and manufacturing process, including consumer reviews and in-service information feeds from devices in the field

One possible way to avoid

DIGITAL AMNESIA

is by keeping everything—or nearly everything.



not only the instructions for producing the object but also the insights and decision-making process that created the design and the journey of the product from concept to disposal. Such information may create new opportunities to gain insights and accelerate adoption of new ideas.

Thinking of manufacturing this way changes the value proposition for data. Too often today, organizations place value only on highly rarified and filtered information: the Six Sigma data, the electronic build record, geometric dimensioning and tolerancing of the design specification, and the final digital thread. With data-driven innovation, however, unexpected value may come from anywhere, including data sets that capture normal, day-to-day activity. As information storage improves to the point where it is limitless and free, we can consider the implications of capturing every bit of data produced at every point in a product's lifecycle.

That might seem like overkill, but I think engineering and manufacturing enterprises will recognize that all product lifecycle data is potentially valuable, from requirements gathering to performance monitoring and evaluations of in-service product. Capturing the exploratory process at the beginning of the design process will have an important impact on designers, especially when all the wandering branches of thought throughout the design process can be replayed. Access to that kind of information will turn design into a process of discovery and exploration akin to data-driven science.

The real cost of accessing this information is in human attention.

Consider the problem currently challenging the community concerning the acceptance and adoption of new materials and manufacturing processes, especially those based on additive manufacturing. Our existing science behind subtractive manufacturing is based on more than a century of work in metallurgy and mechanical engineering. We can accurately predict, for example, the properties of rolled Ti-6Al-4V with great confidence, but additive processes enable us to create entirely new, complex, non-homogeneous materials. Methods for predicting properties of what we can now create cannot possibly be based on exhaustive analysis of physical test coupons. Rather, reaching new material configurations and synthesizing material architectures to satisfy complex, multi-physics design constraints will require new data and computational methods.

Old notions of qualification and certification will yield to those based on data analytics using the sparse sampling of new physical tests, as well as computational observations and predictions. The implications for design tools are significant, and this is but one of the topics being explored in DARPA's new Transformation Design, or TRADES, program.

That vision of design requires engineers, designers, and manufacturers to reconsider some long-held beliefs about best practices. Before we can embrace the future, we first have to let go of the past.

For instance, we must resist the natural engineering tendency to insist that this cornucopia of data should be stored in enduring, standard, digital formats. That is an unachievable, even quixotic goal. Rather than wait for one uber-format, we should harness exponential progress in computing and information science. Continuing improvements in tools for data management, information extraction, schema learning, and machine reading will change the entire calculus around data standards and formats.

We will also have to learn to recognize great ideas wherever they emerge. Tools that can extract the expertise embedded in the data will enhance everyone's capabilities and melt the distinction between expert and amateur. In areas such as finance, commerce, and entertainment, well-nourished computational engines already are yielding astonishing, previously unforeseeable results and unleashing creativity, turning novices into gurus.

Engineering enterprises and their design-to-manufacture partners also must start to exchange ideas and learn from each other. Too much data today is undiscovered or lost to memory, the net asset value of data is lessened, and learning from past innovations becomes difficult, if not impossible. By sharing ideas and information, designers and manufacturers can avoid wasteful design iterations and higher costs while retaining relevant data for post-production activities and new product development.

One possible way to avoid digital amnesia is by keeping everything—or nearly everything. There is real value in the informal information found in engineering notebooks, and there are certainly discoveries that have been lost due to fading memories or deteriorated and destroyed records. Today, with the price of storage plunging, the real cost of accessing this information is in human attention. It is not economically viable, or even feasible, to sift through and accurately curate the data; this must become a job for software. Technologies for doing this automatically are improving exponentially.

We also have to find ways to accept data that is uncertain or potentially unreliable. The key here is to develop the means to quantify the data's utility and track its use (sometimes called provenance or pedigree) over time. This sort of thing is done now with product reviews and scientific datasets. Data's value is not static, and its utility is often discovered or even enhanced over time. The highest quality data can be tracked and tagged.

In time, sufficiently advanced computer-assisted engineering design should enable a kind of digital serendipity. Instead of simply interrogating data to find something specific, we can engage in a more free-form exploration of data that enables discovering the unexpected. That new approach can be viewed as a form of intelligence amplification, where the strategic application of computation and data becomes the key to accelerate scientific discovery, design creativity, engineering proficiency, and other forms of innovation. That sort of computationally enhanced thinking can lead us on a journey that one could not take by oneself; it can lead down paths never imagined or otherwise discoverable.

Far more significant than simply increasing the pace of design iterations, designs themselves can be greatly improved, thereby greatly reducing the rate at which a better design is achieved.

Establishing the rightful and more prominent role of data in engineering, creating a rich and accessible data substrate, and nurturing computational engines will remove unnecessary entry conditions to the design process. As with software development, the future belongs to the designer, not the technician. A larger and more diverse population will be able to participate in the creative process.

But we aren't there yet. The current state of manufacturing design is grounded in 200-plus years of history during which that activity was viewed through the eyes of physics and mathematics, or even as branches of applied physics and mathematics. Computation and data are newcomers. Rethinking existing problems around the notion of algorithm and the abundance of data will require us to discard many long-held beliefs in the light of the new paradigm created by computing machines.

Computer scientists are developing solutions that might enable this transition, but enterprises need to commit to underlying processes and cultures that will truly affect change. The results will be worth it. Let's get started. **ME**

WILLIAM REGLI is deputy director of the Defense Advanced Research Projects Agency's Defense Sciences Office. He may be contacted at regli@darpa.mil.