How data analytics, machine learning, and other advanced technologies are accelerating science

Transforming R&D

Digital innovation in the pharmaceuticals and chemicals industries

Preface

This report, "Transforming R&D: Digital innovation in the pharmaceuticals and chemicals industries," is an MIT Technology Review Insights white paper sponsored by PerkinElmer's Informatics division. It explores how scientists, naturally leading adopters of new experimentation techniques, are harnessing digital technology to rethink their craft, navigate shifting regulatory landscapes, and innovate faster and more effectively than ever before. This white paper is informed by interviews, conducted in October and November 2020, with R&D experts across eight leading global scientific companies. It explores digital transformation strategies in the pharmaceutical and chemical industries–two sectors that, combined, shape almost every aspect of our lives and our future. Adam Green was the author of this report, Claire Beatty was the editor, and Nicola Crepaldi was the producer.

We would like to thank the following experts for sharing their time and expertise.

Dimitris K. Agrafiotis, Chief Information Officer, Novartis Institutes for BioMedical Research (NIBR), United States

Martin Clough, Head of Crop Protection R&D Technology and Digital Integration, Syngenta, Switzerland

Thomas Jung, Head of R&D IT, Syngenta, Switzerland

Yoshito Nakanishi, Head, Digital Strategy Department, Chugai Pharmaceutical, Japan

Bryn Roberts, Global Head of Operations, Pharmaceutical Research and Early Development, Roche, Switzerland

Carol Rohl, Executive Director, Global Research IT and Informatics, Merck, United States

Anthony Rowe, Senior Director and Business Technology Leader, Janssen Pharmaceutical Companies of Johnson & Johnson, United States

Nadeem Sarwar, President, Eisai Center for Genetics Guided Dementia Discovery, United States

Brian Standen, Head of Digitalization in Materials and Chemicals R&D, North America, BASF, United States

Richard Trethewey, Vice President Digitalization in Bioscience R&D, BASF, Germany

CONTENTS

01 Executive summary

Scientific research and development (R&D) has powered human progress for two centuries, but the imperative to innovate keeps intensifying. Chronic diseases like cancer and dementia are a growing burden as societies age, with few treatments despite billions of dollars of R&D investment. As covid-19 shows, new diseases are increasingly likely to emerge as we encroach further into natural ecosystems. Climate and ecological challenges require breakthroughs in everything from the food system and transport to materials science and energy.

As in other sectors, science-based industries such as pharmaceuticals and chemicals have looked to technology to rethink their processes and become more competitive. This digital transformation is reshaping every aspect of R&D, from designing molecules and drugs to facilitating predictive analytics that can sometimes replace "wet" lab experiments entirely. It is, in effect, accelerating science.

Yet while scientists are natural adopters of new technology, pharma and chemicals companies are not immune to the challenges that plague other industries in harnessing digital innovation effectively. Who should lead the process, and where should investment focus? What organizational adaptations are needed? What should the talent pool look like? What cultural changes might be required? This white paper, drawing from wide-ranging interviews in leading life-sciences and chemical companies, draws out critical themes from their journeys thus far. The key findings are:

- **Digital technology is reshaping every aspect of R&D, from design to experimentation**. Researchers are using advanced modeling to design new molecules, identify the strongest candidates prior to costly experiments, and generate richer data, which help lower the costs of drugs and develop environmentally-friendly agriculture.

- **Science-based companies are placing big bets on digital transformation with executive sponsorship and major infrastructure investments.** Leading pharmaceutical and chemicals companies have organization-wide strategies that include investments in data architecture, informatics, artificial intelligence, and, eventually, quantum computing. Scientists are naturally intrigued by new data and technology, and our experts cite the importance of encouraging digital innovation where it naturally flourishes, and then creating opportunities for those approaches to scale. Software industry practices, such as agile development, hackathons, and product-based technology roadmaps, are helping companies nurture creativity.

- **Data governance is essential to leverage digital innovation.** Today's breakthrough computing tools depend on rich and accurate datasets; without those, innovations like machine learning have disappointed in the past. Science-based organizations are focusing on data cleansing, categorization, and lifecycle management to ensure that data is FAIR (findable, accessible, interoperable, and reusable), has detailed metadata, is usable through tools like visualization, and that storage costs don't spiral.

- **Digital transformation disrupts tasks, jobs, and even corporate culture, requiring careful management.** Technology is not just augmenting R&D but fundamentally changing it, automating tasks just as it creates new functions. Positively, this is leading to larger, more diverse talent pools and a blurring of boundaries between disciplines—particularly since covid-19 has brought once laboratory-bound scientists into leadership forums and new forms of communication have allowed R&D teams to participate more in business strategy. But any change in roles and responsibilities will naturally meet resistance. R&D leaders cite the importance of creating excitement, celebrating wins, and providing time to adapt.

02

Digital drivers

The pharmaceuticals and chemicals industries are no strangers to technology and data, compared to sectors such as retail, whose most fundamental "bricks and mortar" model has been upended by e-commerce. "The entire business of drug discovery is about generating data and we've been doing that for 150 years," says Anthony Rowe, senior director and business technology leader at the **Janssen Pharmaceutical Companies** of Johnson & Johnson. Bryn Roberts, global head of operations at **Roche Pharmaceutical Research and Early Development**, makes a similar argument: "Digital transformation has entered our vocabulary in the last two or three years, but we've been doing it for decades."

In the chemical sector too, heavy-duty computational, statistical, and mathematical tools are critical to modeling, simulating, and designing the structures and behaviors of molecules and solids. What is changing, experts attest, is the power of digital technology to fundamentally alter every step in the R&D journey, from how they design new formulations and drugs to the ways they are tested, trialed, and manufactured.

Push and pull

Both push and pull factors fuel digital transformation in scientific R&D. Push factors include the sustainability imperative, especially in the chemicals sector, the industry behind the fertilizers and pesticides that made it possible to feed a growing global population. This has, however, come with a significant environmental toll, and the food

"Digital transformation has entered our vocabulary in the last two or three years, but we've been doing it for decades."

Bryn Roberts
Global Head of Operations, Pharmaceutical Research and Early Development
Roche

sector is a significant producer of greenhouse gas emissions. Chemicals companies are also essential to the petrochemical sector that is under increasing pressure to clean up its environmental footprint.

Tightening government regulation on emissions and climate targets, combined with public disquiet at the scale of the environmental crisis, are pressuring the industry to keep growing output while lowering the use of conventional inputs. The European Union's Green Deal seeks a 50% reduction in pesticides by 2030, a 25% reduction in fertilizer use, and a 25% increase in organic farming. "You've got to do something pretty different to be able to deliver food at a competitive price [in this framework]," says Martin Clough, head of crop protection R&D technology and digital integration, **Syngenta**. "It is a brave vision, but it hasn't yet

got a plan behind it. Digital has got to be part of this." A second pressure is the elimination of animal testing, central to R&D in pharmaceutical research. The US Environmental Protection Agency, for instance, has set a goal of reducing funding for mammal studies by 2035; computational modeling and testing on cells could offer a more humane alternative.[1]

For the pharmaceutical sector, regulatory and social pressures are primarily linked to the price of medicines. As populations age, the cost of treating chronic conditions is putting more financial cost on both public health resources and, in privatized or hybrid markets such as the US, on citizens. One driver of cost is the clinical trial process, which consumes up to 50% of the average $2 billion sum needed to bring a new product to market.[2,3] Returns on R&D have also been steadily falling over the past decade, from 10.1% in 2010 to 1.9% by 2018 (Figure 2). By doing more prediction, modeling, and "dry" science, researchers could improve the quality of new candidates they bring to a trial in the first place.

"If you look at the top 10 companies by market capitalization, twenty years ago, they were mixed industry companies; now they are tech and cloud companies. How do we navigate through this and compete?"

Richard Trethewey
VP of Digitalization in R&D
BASF

Figure 1: Costs of US clinical trials by therapeutic area and phase (US, $ million)
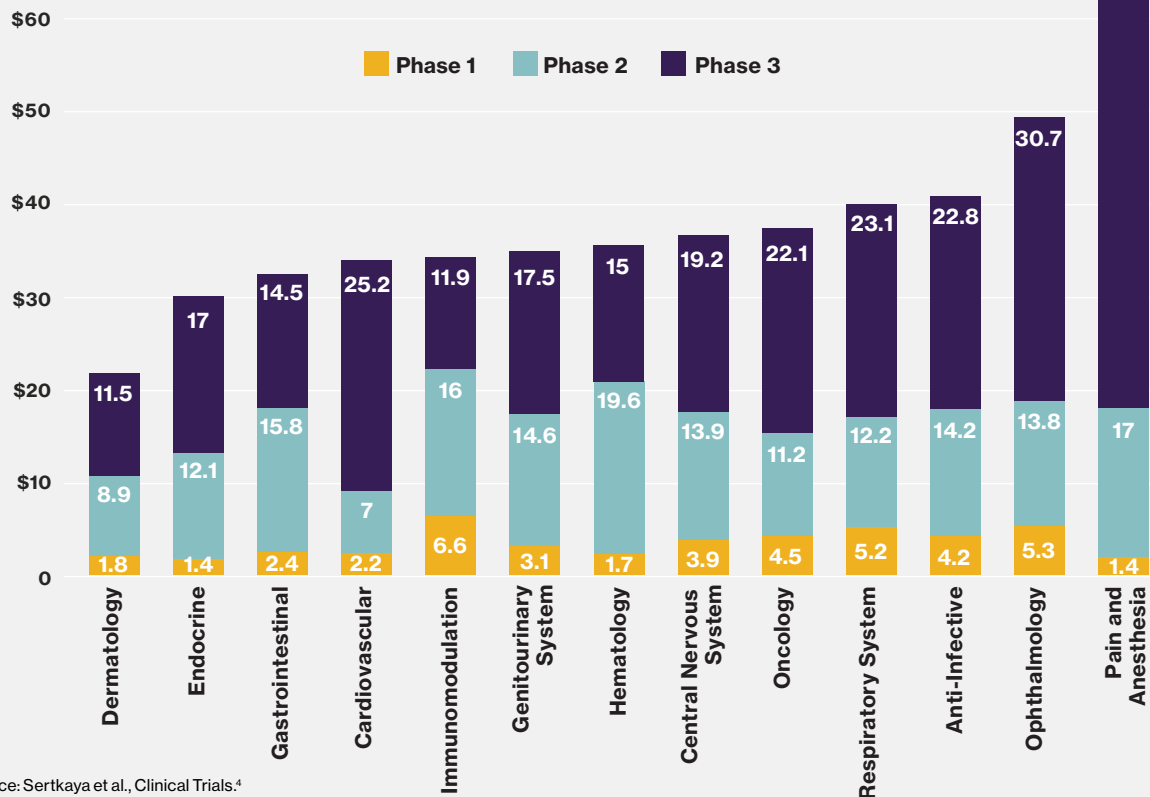
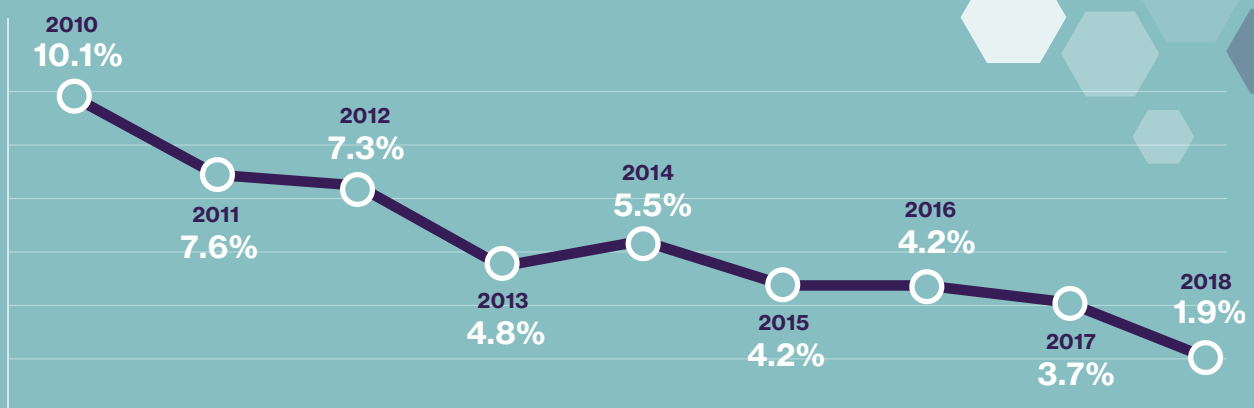Source: Sertkaya et al., Clinical Trials.[4]

Digital innovation will also facilitate smoother interactions with regulators, such as speeding up the work of regulatory filings and turnaround cycles, says Roberts. "Full electronic dossier submission [to regulators] is coming in and manual review processes will be gradually replaced by automation. I could see a situation where we have much more rapid review and approval cycles." Regulators themselves are also taking advantage of digital technologies for review and data analysis, he adds. This could lead to regulators asking companies for more data, meaning the latter will need to have their house in order.

Finally, both industries face a push of competitive pressure. Big tech companies are continually testing the boundaries of industries, moving into everything from finance to health care. Their huge financial resources, data capabilities, and expertise make it possible for them to move into sectors that once had high barriers to entry. "If you look at the top 10 companies by market capitalization, 20 years ago, they were mixed-industry companies; now they are tech and cloud companies," says Richard Trethewey, vice president of digitalization in R&D at **BASF**. "How do we navigate through this and compete?" Even startups could be competitors—note the wave of well-funded deep-tech synthetic biology companies tinkering with microbes to create new classes of materials, from animal feed and fuels to industrial coatings.[6,7]

Returns on R&D have been steadily falling over the last decade, from 10.1% in 2010 to 1.9% by 2018, according to research by Deloitte

Figure 2: Returns to R&D in pharmaceuticals

2010 **10.1%**
2011 **7.6%**
2012 **7.3%**
2013 **4.8%**
2014 **5.5%**
2015 **4.2%**
2016 **4.2%**
2017 **3.7%**
2018 **1.9%**

Source: Deloitte, 2018[5]

03 Doing better science

Digital transformation is helping science-based companies respond to both the pressures and opportunities they are facing. Artificial intelligence and machine learning, fueled by richer datasets like genomics, are opening new horizons in drug discovery by allowing researchers to leverage human biology rather than rely on animal models. Sensors and the internet of things allow researchers to monitor the effectiveness of new therapies or compounds at greater resolution. Capabilities once limited to large laboratories, such as cryogenic electron microscopy, are becoming more widely available. At the frontier, innovations in quantum computing could enable the design of new materials.

One overarching application is improving the quality of new products before they enter the experiment and trial process. Analytics, simulation, prediction, and modeling show researchers more about how a new innovation might fare in the real world. This eliminates candidates that might once have incurred costs in a failed pilot or trial due to weaknesses that researchers were unable to spot in an earlier phase.

The right study, faster

Chemicals researchers, for instance, could identify the most effective molecular designs or structures before having to synthesize molecules in laboratory settings. "If digital innovation helps us get to the right study faster, it helps us shave off time and cost," says Clough at Syngenta. Roberts refers to it as using modeling to increase the "probability of technical success," or PTS in industry lexicon. Yoshito Nakanishi, head of the digital strategy department at **Chugai** Pharmaceutical, says that

AI and digital "can help us glean the best compound in the early stage, so we can skip over these very long research processes. We use AI to select or optimize the antibodies and evaluate efficacy or toxicity in the preclinical stage."

Brian Standen, global head of next-generation computing at BASF, sees digital technology and experimental science as two interlocking gears. "The first gear is the work you can now do digitally, in silico, which drives laboratory experimentation in the lab, which then feeds back. Digitalization explores the 'space of interest' for discovery, to guide experimentalists to do more focused experiments, and they come back with the data to refine that process further."

Digital innovation can then improve the quality of experiments once underway, from optimizing design and protocols to generating better data about effectiveness. AI can improve patient-trial matching by crunching large datasets like electronic health records, genomic data, medical literature, and trial databases. Inadequate enrollment of patients in trials—too small a sample, too high a drop-off rate, or suboptimal cohort selection, such as failing to exclude individuals using another medicine that could interfere with the tested one—is a major cause of trial failure.[8]

Once underway, advanced analytics can enable adaptive clinical trials, where researchers modify a live trial, such as by refining the sample size, changing the allocation of patients to control versus trial arms, and stopping a trial early based on likely success or failure.[9] Digital technologies like wearables, implants, and digestibles can show how a therapy is working in richer detail via digital biomarkers,

according to Rowe at Janssen Pharmaceutical Companies. "You want to understand fluctuations in a disease and how effective a therapy is [at a granular level] ... if you can tell how someone is sleeping, how cognitively alert they are, their fatigue, their mobility, these things give a much better indication of whether a drug is helping."

Chugai Pharmaceutical's Nakanishi makes the same point, describing current research efforts to use biomarkers to track symptoms of endometriosis, a painful uterine disorder. "Usually when we would like to evaluate pain, we have a number scale from 1 to 10, asking 'what is your level of pain today?'", says Nakanishi. "It's not a truly accurate way to evaluate pain. Using digital biomarkers, we can get vital data like heart rate or dermal activity, and if you combine this data with AI, we can estimate a pain index and better understand how painful it is right now for the patient and visualize this."

Digital biomarkers could eventually enable the decentralization of clinical trials, reducing the need for patients to attend some or all in-person consultations.[10] This could further expand trial enrollment by allowing participation of people who live in more remote or poorer areas, which tend to be further from clinical research centers. It could also allow more racial and ethnic minority participants for whom location and logistics are currently more of a constraint to trial participation.[11]

Digital biomarkers could allow preventive medicine in conditions like cardiovascular disease or multiple sclerosis, where acute events like strokes or demyelination symptoms can have long-term consequences. "With diseases like multiple sclerosis, you can see someone trending towards an adverse health event, giving you a window to intervene, such as to get them a dose of steroids," says Rowe. One current trial led by Johnson & Johnson and Apple is exploring if a smartphone app could detect atrial fibrillation, a heart rhythm irregularity that is a precursor to strokes.[12] If successful, this could support a new wave of R&D into preventive therapies to intervene before an adverse event happens in the first place.

Breaking ground

Lowering the costs of trials and improving experimental data are worthy ends in themselves, but companies also see an opportunity to use digital technology to do entirely new science. In health care, for instance, the quality of data available to researchers is helping the development of therapeutic areas like immunotherapy, which alters the human immune system to better fight diseases like cancer. This is a far cry from current therapies like radiotherapy and chemotherapy, which have changed little in decades, bring debilitating side effects, and are often palliative rather than curative—even as cancer prevalence is set to increase significantly over the coming decades (see Figure 3).

Figure 3: Estimated number of new cases from 2020 to 2040, both sexes, age 0-85+ (All cancers)

19.3m

30.2m

■ 1,000,000

□ Demographic changes

Source: World Health Organization International Agency for Research on Cancer, 2020[13]

"We have an unprecedented opportunity to understand human biology in a much higher resolution that can drive new therapies and new modalities, from antibodies to small molecules to everything in between," says Carol Rohl, executive director of global research IT and informatics at **Merck**. "And product now is not just the therapy or device, but the data around it," she adds, emphasizing the importance of data that show how a therapy's impact differs across groups.

Genomic data is fueling much R&D optimism in the pharmaceutical industry today. Historically, drug development relied on mouse models with a hope that findings would translate to humans, leading to high failure rates. Rapid growth in genome sequencing technology, and falling costs, allow researchers to design therapies based on human biology. National genome data mapping initiatives, which kicked off in 14 countries between 2013 and 2019 with an estimated $4 billion of investment, are research treasure troves.[14]

"Therapeutic targets anchored in robust human biology data, especially human genetic data, are much more likely to be successful," says Nadeem Sarwar, president of the **Eisai Center for Genetics Guided Dementia Discovery**. "Our ability to study human genetics at unprecedented scale, speed, and detail has enabled a brand-new paradigm for drug discovery and development." Sarwar believes genomic data could improve understanding of genetic differences between racial and ethnic groups. "The therapeutic insights you get from genomic data are different depending on which population you study. The vast majority of our available data comes from populations of predominantly European continental ancestry, and there's a strong push to think about identifying previously underserved populations and identify novel genetic associations that may be especially informative for therapeutic discovery."

Real-world evidence data gathered from outside of clinical research settings, like electronic medical records, claims and billing data, and disease registries, is another data resource opening new drug discovery prospects, especially when combined with genomics, says Sarwar. "Medical records provide opportunities to do genotype-phenotype studies at unprecedented scale. For example, countries like Scotland and Estonia have exceptional real-world data, and electronic health records allow you to have very large amounts of health data that have already been collected, and be able to relate them to genotypic data that you may collect in new studies."

"We have an unprecedented opportunity to understand human biology in a much higher resolution that can drive new therapies and new modalities, from antibodies to small molecules to everything in between."

Carol Rohl
Executive Director, Global Research IT and Informatics
Merck

As important as new data sources is the rise in collaboration and data-sharing initiatives, creating shared assets that benefit the entire community. The European Union is a lead player; its Open Science initiative seeks to give the continent's 1.7 million researchers and 70 million science and technology professionals access to world-leading data management tools and high-performance computing.[15,16]

R&D experts expect quantum computing to open up new materials possibilities in niches like batteries, semiconductors, and superconductors, as well as more complex assembled products like composites and opto-electronic devices.

Researchers in specific disease categories are federating and pooling data that tends to fragment and silo. Project Data Sphere, an oncology initiative, aggregates data from pharmaceutical companies, medical centers, and government organizations. The Bill & Melinda Gates Foundation is also investing in data sharing in dementia, an area of growing interest to the organization.[17]

Evolution, accelerated

In the chemicals sector, digitalization's use cases include designing and modeling new materials and composites. Chemicals innovation is woven into a dizzying number of everyday products, from smartphones and fiberoptic network cables to autonomous vehicles; 90% of everyday products contain chemicals.[18,19]

"Digitalization enables us to better predict the properties of a material, the performance of an active molecule, or its toxicological effect," says Trethewey at BASF. "The frontier it really opens up is going for products which you wouldn't necessarily have thought of through classical methods," he adds, likening it to route suggestions on Google Maps that a traveler may not have considered.

Santa Monica-based Provivi, a BASF venture-backed startup, is developing artificially-synthesized pheromones–the wind-borne compounds that insects use to communicate–to control populations, such as by disrupting the mating cycles of crop-harming insects. One of the most hyped tech sectors of today, alternative proteins, is also at the intersection of chemistry, biology,

and sustainability. The meat-like burgers of Impossible Foods are based on the fermenting of engineered microbes–the same ideas are driving the sustainable materials segment, like the bio-engineered silk from Bolt Threads.

R&D experts expect quantum computing to open up new materials possibilities in niches like batteries, semiconductors, magnets, and superconductors, as well as more complex assembled products like composites, effect pigments, and opto-electronic devices, which would benefit from the simulation, AI, and optimization capabilities of quantum power.[20] BASF's recent investment in quantum computing startup Zapata, a Harvard University spinout, promises to help it shorten development cycles.

"We have certain limits in terms of the computational power available to us today, so we have to take more coarse-grained approaches to understanding structures of molecules, structures of proteins," says Standen. "Quantum computing can allow deeper molecular modeling." While a range of industries could gain from quantum computing, including finance, pharmaceuticals, automotive, and aerospace, the chemicals sector might be among the first to reap the fruits. "Chemistry could be impacted far sooner than pharmaceuticals or banking, partially because of the number of qubits [the units of calculation] needed to have the impact," he says. "Quantum modeling is more near-term for us than the more complex models you would try to organize for pharmaceuticals."

Strategies and best practices for digitalizing R&D

The power of digitalization to accelerate R&D is largely thanks to performative improvements in the "tech stack:" more data leads to better algorithms; cheaper, better sensors capture more information; more powerful computers drive better predictive modeling and simulation.

But technology is never the whole story. It must be deployed to solve the right problem, and sometimes tried-and-tested approaches work just as well or better. Powerful tools can derail without human guidance and iteration, as with "model drift" that bedevils some AI systems over time. Researchers need to be enabled to explore new technologies, without having them pushed on them. The talent pool may need to change, and so might organizational structures. R&D experts have already learned useful lessons from their digital transformation journeys thus far.

Balancing top-down direction with grassroots energy

Scientific R&D has often been monolithic in nature, due to the capital and resources needed to conduct research and the enterprise planning to support it. This has led to fundamental advances in science, such as the Manhattan Project, the first "big science" initiative that led to the development of the atom bomb that ended the second world war.

Today, digital transformation is more multidirectional. Top-down support is essential to drive the big investments

In leading organizations, digital transformation is multidirectional. There is bottom-up innovation to scale pockets of excellence and top-down support to drive the necessary cloud and data architecture investments.

in cloud computing and data architecture required; At BASF, for instance, digitalization is one of six strategic action areas defined in the corporate strategy. Its development was guided by the CEO who is also the company's CTO. But companies also need to harness grassroots ideas and energy, particularly as pockets of innovation and excellence emerge. Leading organizations are doing this by creating platforms and infrastructure that allow these leading practices to scale.

R&D experts cite the rise of agile working principles, popularized by the software sector, with an emphasis on iteration, feedback, sharing, and scaling-up, rather than top-down reform. "You don't want to be transforming forever or thinking in terms of huge change projects," says Roberts at Roche. "You need to have a continuous-improvement mindset." Rowe makes a similar point. "There is no overarching plan to transform digitally. Instead, there is a constant digital transformation process."

Syngenta takes a "leader-led" approach, that leverages grassroots enthusiasts and provides communications

"There is plenty of low-hanging fruit in the early stages of applying new digital tools and processes, which can get the conversation started. You can take easy-win categories of problems to help a scientist see how digital can change what they do."

Richard Trethewey
VP of Digitalization in R&D
BASF

platforms that help those efforts and ideas percolate. "We're going into the organization, rewarding innovation, raising the level of consciousness about innovation, and driving it forward through proof of concepts," says Clough.

As well as looking at how scientific teams are using digital technology, R&D leaders are also encouraging experimentation, without forcing it. "This needs to be about 'citizen development,' allowing people to empower themselves with digital tools and get interested in them independently," says Standen at BASF. "There is plenty of low-hanging fruit in the early stages of applying new digital tools and processes, which can get the conversation started. You can take easy-win categories of problems to help a scientist see how digital can change what they do." Trethewey concurs. "You can't come in and just say, 'use this tool because your ability to do chemical formulation is going to increase 100%' ... you need to include people and understand their situation and help them come along, have them start articulating how the digital tool will help them."

Toward product organizations

Where centralization and planning come in useful is in creating structure for innovation and ensuring that efforts are well-orchestrated. "We have to find that balance between creative diversity and constant evolution [on the one hand], with doing things professionally and long term," says Trethewey. To achieve that, BASF introduced a product-based approach in some important areas by appointing product managers with end-to-end accountability for the success and adoption of the relevant technologies. In other domains, BASF looks for more creative innovation and even replacing the status quo. "You need a framework to operate in, with defined areas not to be touched and other areas which are a sandbox."

At **Novartis Institutes for BioMedical Research** (NIBR), Chief Information Officer Dimitris Agrafiotis is leading a major program to build a product-based organization. "Historically, IT or informatics groups would build a system, deliver it, and move on to the next thing. There was no discipline about building technologies that would grow and blossom after their initial release." NIBR's systems have now been reorganized into product lines and individual products, each with roadmaps that orchestrate digital innovation in synchrony. "You don't want to end up with a hodgepodge of systems that don't talk to each other," he says.

Last but not least, there might be a need to create new structures to finance digital innovations that could otherwise be underinvested in, according to Chugai Pharmaceutical's Nakanishi. "Researchers do not have budget for digital activity specifically—their priority is their

activities right now." The digital department that Nakanishi leads is working to change mindsets, so that R&D teams see digital capability as an investment in the future, as well as making additional budgetary resources available that they can pull on for appropriate initiatives.

Data and infrastructure

Digital transformation is fueled by data, but data do not proliferate neatly. The more of it that gets produced, the more challenges emerge around infrastructure, processing, cleaning, wrangling, and sharing. Experts say that investing in high-quality data is the first step toward using data for product development and innovation.

Rowe likens it to Google's efforts to build self-driving cars. "They didn't start with cars. They started with Google Maps and built this amazing resource. They get intermediate value from that, but maps were a step on the way to the self-driving car." AI and machine learning only became useful when the company invested in producing more data by "orders of magnitude."

Then comes architecture and analytics. "The scale, size, volume, and complexity of the data we now generate has required real innovation in the mathematics and analytics of how we store, organize, integrate, analyze, and interpret these data," says Sarwar at Eisai. Roberts at Roche says the company cleaves to the FAIR acronym—findable, accessible, interoperable, and reusable—as its guiding strategy for managing data.

For Syngenta, one data governance focus is "immortalization." "Like many industries, we have historical data typically generated by an individual to answer a question, and once it's answered, it's buried, often without any metadata around it," says Thomas Jung, head of R&D IT. "You don't know how it was generated, who generated it, what the experimental conditions were, and so on. When you then go back and query those data, they're pretty well worthless unless you still have a warm body who can put it in context." Data can be made ageless by furnishing it with enough metadata to give it context, and leveraging data lakes to ensure data, and the descriptors around it, are findable, he says.

Rowe advocates a balanced data management strategy to avoid pushing the data usage problem downstream. Data lakes, for instance, can end up forcing data scientists to become data engineers. "The real trick is to find a happy medium where you process data to the point

Digital transformation is fueled by data, but data do not proliferate neatly. The more of it that gets produced, the more challenges emerge around infrastructure, processing, cleaning, wrangling, and sharing.

where it could be useful, but not overly constrained or processed."

He likens it to gardening where users can "prune and trim, look at where they want to grow a new flower beds and where to mow the lawn. The ideal stack is some form of data lake that is well-managed, with an indexing engine that links data together. You need a balance between control, complexity, and autonomy."

This is where more purposeful organization-wide redesigns might be necessary. Some companies have formed data management groups. Novartis, for instance, created a dedicated unit focusing on architecture, data flow, and ontology—making sure data is annotated, classified, and transformed—and data lifecycle management. "As mundane as it sounds, if we didn't move data from primary to secondary to tertiary data storage, our budget would have been wiped off paying for storage," says Agrafiotis.

PerkinElmer Informatics perspective: critical requirements to achieving digital transformation

Where there is a will there is a way. The importance of achieving digital transformation is now widely recognized, so there will be progress, and technology has improved sufficiently such that digital transformation in the context of scientific research is feasible. However, the challenges remain formidable, both at the level of change management and technical implementation.

Scientists are results-oriented and impatient. They are always asking the questions—how can I do my work faster, and what do I need to do next to achieve my goals? They are also much more strongly influenced by evidence they can observe for themselves instead of logical conjecture. The very same traits that can help them in science can hurt them in the context of digital transformation because, in IT parlance, it is frequently the nonfunctional requirements—the ones you can't see or easily test—that determine the success or failure of an initiative, and this is definitely true of digital transformation initiatives.

From a PerkinElmer Informatics view, two nonfunctional requirements lie at the heart of successful digital transformation projects, and they go hand-in-hand:
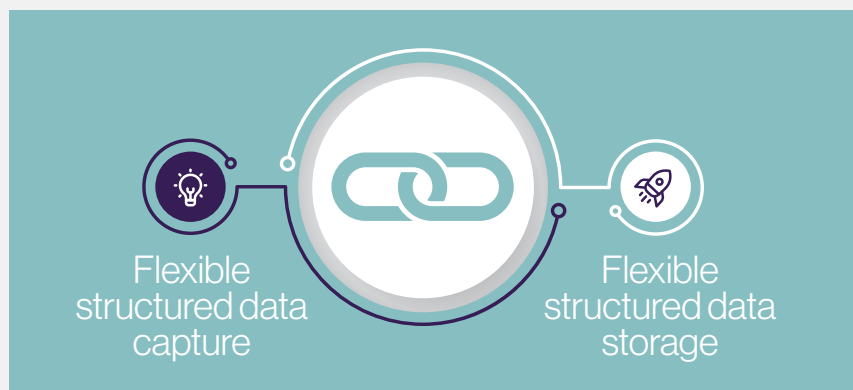
As scientists progress their projects, they acquire data from increasingly complex scientific instruments, process raw output into human-readable results, and then try to collate and align this output into datasets used for decision-making purposes. Core to digital transformation is the notion that these datasets are easily, if not automatically, assembled and ready for more sophisticated analysis. However, in almost every case, digital transformation initiatives in science fail because experimental data is too variable to be structured easily and data storage systems are not adaptable enough to store the data effectively even if it is well-organized.

We urge our customers to look very closely at these nonfunctional requirements before driving forward with digital transformation projects in science. Otherwise, they will fail. We have found that putting a priority on flexible, structured data capture that

flows data naturally from scientific instruments into structured output coupled with flexible, structured storage mechanisms is the only way to achieve the benefits of digital transformation in the context of enterprise science. Achieving these objectives is no mean feat. It requires putting some key functional requirements on the back burner and attaining fluency in a range of new but immature and challenging technologies. By prioritizing these two critical non-functional requirements and patiently understanding and embracing the new technology stacks that have enabled platforms like Facebook and Salesforce, it is possible to achieve the goal of making your data work for you. However, if these twin nonfunctional requirements are not put at the center of the digital journey, you will be on a journey to somewhere, but not where you want to go.

Flexible structured data capture

Flexible structured data storage

05

Digital innovation:
A new world of work

The digital transformation journey is not just about new technology and data, it also requires new ways of working and even a cultural evolution. Huge, age-old conglomerates now talk like startups. BASF, founded in 1865, now runs hackathons to elicit interdisciplinary ideas from across science and data teams and the approach has generated several pieces of new technical infrastructure for the company. "It just works. People love working in that type of environment," says Trethewey.

Jobs in transition

But the cultural transition is not easy. Some researchers can be fearful or resistant to change. Digital pathology, for instance, moves the analysis process from manually reviewing physical slides to inspecting digitized versions on screen to, increasingly, automated image analysis by AI. This can be a good thing for the human pathologist, allowing them to focus on more complex and challenging work, but it fundamentally changes the nature of a job someone may have spent decades doing. "If you've enjoyed a routine job reviewing slides, it is frightening to see that your job might be going away," says Roberts, although he adds that scientists are on the whole excited at the chance to do their job better and faster.

Jung at Syngenta observes that digitization can quickly move from a helpful aid to a perceived threat. "People

have been open-minded about moving from paper to iPad to collect data. Our challenge has been to say, 'but have you thought about not doing the experiment at all, and predicting it instead?' [When you do that] suddenly, you are very close to someone's job. You've gone from helping them to risking their livelihood in a very small step."

Companies must also avoid going to the other extreme and chasing after every new gadget. "At the turn of the century we saw a rise of data mining, machine learning, advanced statistics, and bioinformatics, but we were only partially successful in that era," recalls Rowe. "I am somewhat skeptical this time around, but I do think the tech and science have evolved and there is more readiness to adopt."

It is critical to focus on getting the right match between technology and the available data and not throw out the old for the new. "Machine learning and deep learning might be good, but they are not always the best fit," says Standen. "Sometimes traditional statistical approaches are just as good." Even the most advanced scientific R&D units do not typically generate the kind of data that tech giants might. This means data-hungry analytics tools are more useful in Silicon Valley than they might be in Basel or Cambridge. The key, says Agrafiotis at Novartis, is to work backwards from the overarching goal of the company. "Our mission is to make medicines that help people live longer, better

"We know how to value traditional scientists but we're still trying to figure out how to do that for data scientists. We don't yet have the right career development frameworks for them."

Carol Rohl
Executive Director, Global Research IT and Informatics
Merck

lives, not to develop novel data analytic technologies, and sometimes that can get missed. An outcome-down view, as opposed to a technology-up view, can be important."

Building a digital talent pool
New ways of working also flow from new skillsets and staffing. Digital transformation requires building up the data capabilities of existing teams and bringing in new talent in more specialized domains. Both pharmaceutical and chemical R&D experts interviewed for our report said they were able to attract talent in a competitive market for data scientists. Partly, they are well-paying sectors, albeit not at the level of big tech. But they also have attractive missions. "We have a lot of people that join us saying they want to work on the type of big questions that we're engaged in. They want to work on something connected to the long-term challenges we face," says Trethewey.

Recruitment decision-makers must be as attentive to talent hype as those deciding on technology itself. Experts are cautious about "data jockeys" and hesitant that simply adding a data scientist allows them to break new ground. "They've got to have the science background to be able to work, and to have credibility to work with other scientists," says Jung. "Otherwise, they get marginalized very quickly."

Decision-makers themselves are still grappling with how best to recognize talent for the digital transformation era. "We know how to value traditional scientists but we're still trying to figure out how to do that for data scientists. How do you compare a data scientist with someone who is, say, driving the clinical development of a drug? We don't yet have the right career development frameworks for them," says Rohl at Merck.

There are still talent shortages in the market for the right blend of attributes, leading some companies to develop partnerships to build up the pipeline of future talent. "The digitization of drug discovery and development is a reasonably recent phenomenon, so there is a critical need to train the next generation of medicine-makers to include data and digital sciences as a core drug discovery skillset, akin to how disciplines like biology and chemistry are viewed," says Sarwar at Eisai. The company has partnered with education institutions "from middle school to post-docs," including a collaboration with the Health Data Research UK and the Alan Turing Institute PhD program.[21] Over time, he says, the distinctions between the disciplines of biology, chemistry, and data science will erode.

Covid-19: Democratizing digital?

The pandemic has disrupted scientific R&D significantly, with projects delayed or terminated due to laboratory closures, stay-at-home orders, and the upending of supply chains for equipment and materials. The global collaborative effort to understand and tackle covid-19 has been historic, but this has also drawn resources away from many other important research endeavors.

Yet as with so many other sectors, from the acceleration of e-commerce in hitherto-lagging industries to the normalization of flexible and remote work, there could also be long-term adaptive benefits to the crisis. Productivity and focus can improve for some scientists, says Trethewey, echoing findings in the wider business media indicating that some, although clearly not all, are reaping the benefits of quieter working environments and fewer office distractions.

R&D experts believe the shift to distributed work is having democratizing effects. For instance, conferences whose attendance was once limited by travel budgets and job title are now accessible to more people. BASF's Northeast Research Alliance conference, a collaboration with MIT, Harvard, and the University of Massachusetts Amherst (UMass), previously held two meetings for 100 participants a year. Now as a virtual event, many more staffers are able to join and contribute.

Being creative about conferences has provided greater access to digital science, according to Clough at Syngenta. "Virtual conferencing enables you to bring digital scientists into fora they might not have found easy to join, like leadership meetings. It lowers the barrier of who you can invite to meetings and gives more people a voice." Clough also thinks the pandemic is helping bring researchers out of their offices. "One of the objectives of our digital transformation is to get scientists out of the lab and into the public eye, to get the credit for what they are doing."

As digital scientists gain more recognition, they are encouraged to put their head above the parapet more, creating a feedback loop. This could help them gain recognition for attributes beyond just data and science. "I've seen leadership behaviors from our digital scientists that could have taken longer to get to than pre-covid," says Clough.

Hiring protocols have also become more flexible, opening up the global talent pool. "I've made some hires where location was not an issue and we've tapped into amazing talent," says Agrafiotis at Novartis. Even shutdowns of experiment units have allowed more bandwidth for digital transformation initiatives that might struggle to get attention and time in the normal run of play. "We have been able to get more time from scientists to develop our digital transformation discussions, so that has accelerated our progress," says Roberts.

But R&D leaders also recognize that, just like many other business areas, there is a kind of serendipity that comes with being physically co-located that seems to correlate to innovation and creativity. And that would be very hard to replicate virtually. "What is less clear to me at the moment," says Agrafiotis, "is the impact of virtual workplace on those hallway conversations, the casual encounters, grabbing a beer, having spontaneous conversations. That randomness is not easy to engineer into our virtual experience. You have to program and schedule it."

06 Conclusion

The pharmaceutical and chemicals industries are no strangers to advanced data and technology, with several past hype cycles in artificial intelligence, analytics, and computing serving only to disappoint. Yet R&D leaders do believe the performance levels achieved by today's digital innovations are at a scale and scope that could not just help researchers work more efficiently, but could also open up fundamental new horizons in what science can achieve. Ultimately, these technologies will support pharmaceutical and chemical companies in addressing critical global challenges such as curing diseases and transitioning to a low-carbon economy. This report, based on interviews with R&D leaders in leading corporates across both sectors, reveals the most powerful use cases, strategies, and tactics for success.

- **Digitalization is a critical competitive advantage in the race for scientific discovery and development.** There are multiple compelling push and pull factors that are spurring the advancement of digital transformation programs. The need to meet ambitious targets in emissions and bring down the costs of drug development are also pressures driving R&D teams to explore digital innovations. But they also see an opportunity to do fundamentally new kinds of science thanks to larger datasets, genomics, and new computing paradigms like quantum computing.

- **Digital transformation must be supported with top-down focus and investment as well as encouraged wherever it flourishes at a grassroots level.** R&D leaders emphasize that digital transformation is an evolution, rather than a revolution. While some elements have been slowly emerging over the past decades, the past two to three years have seen the launch of larger, organization-wide transformation strategies at pharmaceutical and chemicals companies. The investment and executive focus channeled into these programs must, however, be mindful not to create the appearance of imposing technologies onto scientists. Just as important are efforts to build on grassroots energy by rewarding and encouraging innovation, and well exploring opportunities to scale new tools. As digital transformation matures, leaders are moving toward product-based organizations to manage the lifecycle of their tools and create coherent and synchronous technology environments.

- **Robust data governance is a foundational capability for high-performance R&D.** Rich, accessible, and shareable data are the fuel on which today's breakthrough analytics and computing tools rely. Without highly accurate categorization, datasets are unusable for the purpose of scientific experiments. This requires pharmaceutical and chemicals R&D teams to focus as much time and resources on their data architecture as on individual digital technology. To make data FAIR—findable, accessible, interoperable, and reusable—companies are developing detailed metadata and governance protocols, and utilizing advanced analytics and data visualization tools.

- **Digitalization is driving new ways of working and a culture shift.** New technology can change jobs and even threaten some roles existentially, leading to fear and conservatism in some quarters. Leaders cite the change in culture as the biggest challenge along the digital transformation journey. Covid-19 has to some extent leveled the playing field between scientists and their digital counterparts by bringing both sets of skills into the same discussions without hierarchy. As digital ambitions grow, science-based companies need to continue bringing in new talent, developing frameworks for evaluating digital skillsets, and providing career paths for R&D data scientists.

About MIT Technology Review Insights

MIT Technology Review Insights is the custom publishing division of *MIT Technology Review,* the world's longest-running technology magazine, backed by the world's foremost technology institution–producing live events and research on the leading technology and business challenges of the day. Insights conducts qualitative and quantitative research and analysis in the US and abroad and publishes a wide variety of content, including articles, reports, infographics, videos, and podcasts. And through its growing MIT Technology Review Global Panel, Insights has unparalleled access to senior-level executives, innovators, and thought leaders worldwide for surveys and in-depth interviews.

From the sponsor

PerkinElmer enables scientists, researchers and clinicians to address their most critical challenges across science and healthcare. With a mission focused on innovating for a healthier world, we deliver unique solutions to serve the diagnostics, life sciences, food and applied markets. We strategically collaborate with customers to enable earlier and more accurate insights supported by deep market knowledge and technical expertise. Our dedicated team of about 14,000 employees worldwide is passionate about helping customers work to create healthier families, improve the quality of life, and sustain the well-being and longevity of people globally.

PerkinElmer®
For the Better

Footnotes
1. "EPA move to phase out animal experiments could mean the end of toxic regulations", The Intercept, 3 July 2019.
2. Harrer et al., "Artificial Intelligence for Clinical Trial Design", Trends in Pharmacological Sciences, August 2019.
3. Franzen et al., "Impact of organ-on-a-chip technology on pharmaceutical R&D costs", Drug Discovery Today, September 2019.
4. Sertkaya et al., "Key cost drivers of pharmaceutical clinical trials in the United States", Clinical Trials, April 2016.
5. Deloitte, "Unlocking R&D productivity: Measuring the return from pharmaceutical innovation 2018". (pdf, p. 4)
6. "SoftBank-Backed Synthetic Biology Startup Raises $300 Million to Build a Better Microbe", Bloomberg, 8 September 2020.
7. "With $300 Million In New Funding, Zymergen Aims To Sustainably Transform The $3 Trillion Chemical and Materials Industry", Forbes, 10 October 2020.
8. Fogel, "Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: A review", Contemporary Clinical Trials Communications, September 2018.
9. Pallmann et al., "Adaptive designs in clinical trials: why use them, and how to run and report them", BMC Medicine, 28 February 2018. Contemporary Clinical Trials Communications, September 2018.
10. Coravos et al., "Developing and adopting safe and effective digital biomarkers to improve patient outcomes", NPJ Digital Medicine, 11 March 2019.
11. "How to stop a lack of diversity undermining clinical trial data", Financial Times, 18 January 2019.
12. "Johnson & Johnson Launches Heartlineä, the First-of-its-Kind, Virtual Study Designed to Explore if a New iPhone App and Apple Watch Can Help Reduce the Risk of Stroke", Johnson & Johnson, 25 February 2020.
13. Estimated number of new cases from 2020 to 2040, Both sexes, age [0-85+], World Health Organization International Agency for Research on Cancer, 2020
14. Stark et al., "Integrating Genomics ainto Healthcare: A Global Responsibility", The American Journal of Human Genetics, 3 January 2019.
15. "Open Science", European Commission.
16. "The European Cloud Initiative", European Commission.
17. "Why I'm digging deep into Alzheimer's", Gates Notes, 13 November 2017.
18. "DuPont launches with a new take on innovation", Chemical & Engineering News, 9 June 2019.
19. "Digital Transformation Initiative: Chemistry and Advanced Materials Industry", World Economic Forum, January 2017. (pdf, p. 6)
20. "The next big thing? Quantum computing's potential impact on chemicals", McKinsey & Company, 12 July 2019.
21. "Wellcome to fund Health Data Research UK and The Alan Turing Institute PhD Programme," Health Data Research UK, 16 September 2019

Illustrations
Illustrations assembled by Scott Shultz Design with elements from Luchelle, nubnamo, and phitabig, Bianca Neve, and davooda, Shutterstock.

MIT Technology Review Insights