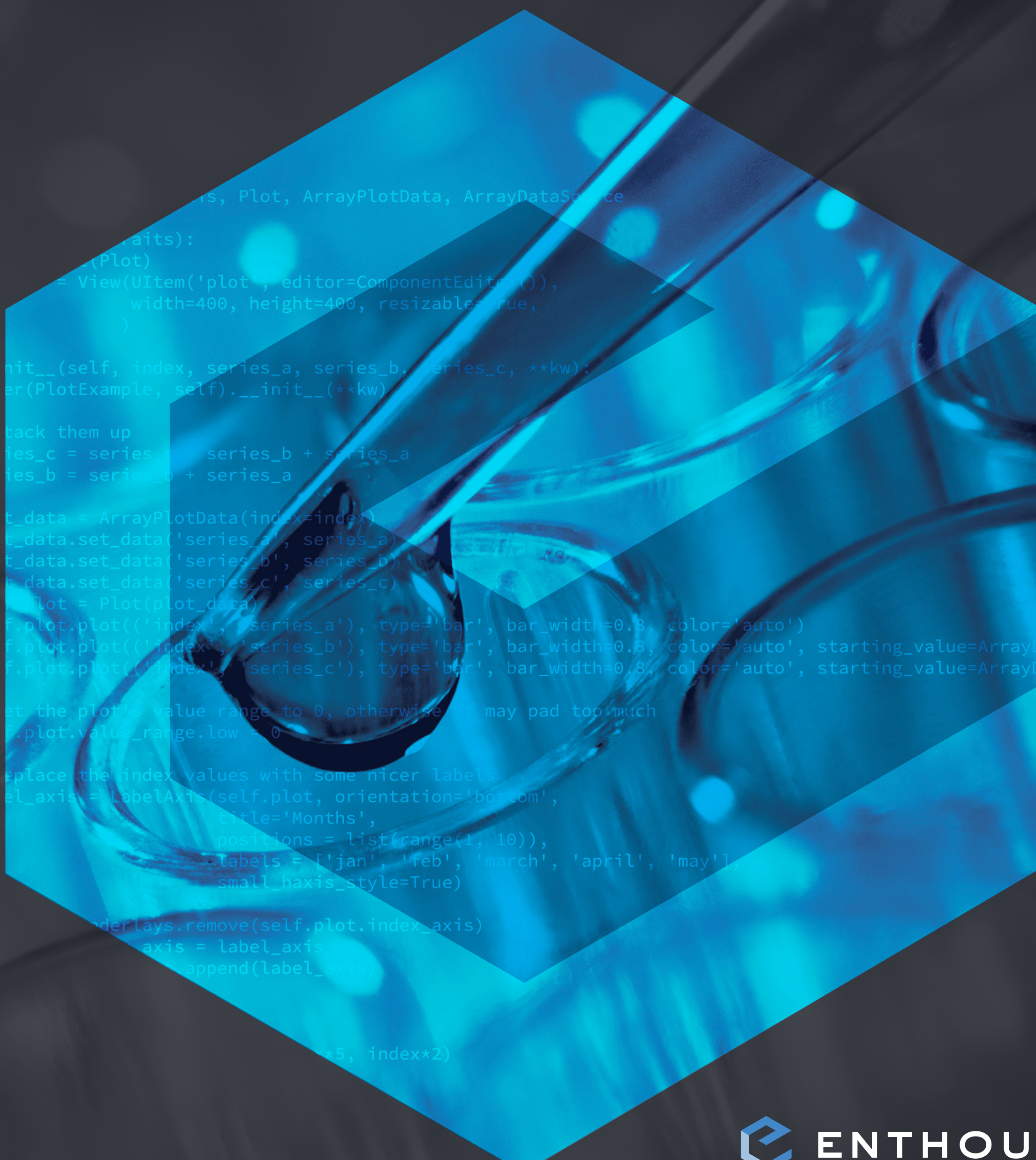
White Paper

The Journey to Digital-centric Chemicals and Materials Laboratories

Digital technologies and the innovation they enable are steadily transforming chemicals and materials labs. This white paper proposes five ‘levels’ of lab digital capability, providing managers and scientists with a mental model to evaluate where they are now, and how to ‘level up’, ultimately transforming lab performance and its impact on the business.



ENTHOUGHT

The Journey to Digital-centric Chemicals and Materials Laboratories

This white paper examines how digital technologies and the innovation they enable are making their way into chemicals and materials labs. It offers five ‘levels’ of a digital capability journey, focused on transformative acceleration of new materials development using the latest digital technologies, including data management, machine learning, simulation, and automation. These levels provide managers and scientists with a mental model for where they are and how to ‘level up,’ ultimately to a point of transforming both lab performance and its impact on the business. This paper draws on over a decade of Enthought experience working with companies worldwide to support transforming their R&D labs through digital technologies, skills, and innovation.

As with any significant new technology arrival, there is a spectrum of adoption by companies, from early ones to laggards, with ‘the chasm’ in between. However, adoption of the rapidly advancing digital capabilities can be ill-defined, with a very different character from that of introducing ‘traditional technologies.’ Incremental improvements resulting from the introduction of new digital technologies can give a false sense of transformative progress as existing workflows show some efficiency gains.

First among the concerns in digital lab transformation are the costs and risks of change, most often in equipment and processes that have been delivering results for years, if not decades. Then there are personnel concerns: transitioning the deep expertise of long-serving senior scientists who are equally concerned with their futures, ensuring that the next generation of scientists have the necessary skills to take full advantage of digital advances, and finally, where to start.

Labs want a smooth transition that delivers value early and continuously. Many R&D leaders acknowledge that transformational change must happen for their business to remain leading or, at a minimum, stay competitive. However, choosing where and how to start remains a challenge.

This paper provides a framework and series of critical steps for achieving transformative results in a dramatically accelerated R&D lab, which will elevate the business it supports in ways they could only dream of.

Five 'Digital Levels' of R&D Labs: There are five levels for a lab to self-evaluate where it is on the digital spectrum, with a focus on transformative acceleration of new materials development using the latest digital technologies. From there, current initiatives can be evaluated and new ones planned to level up in both capability and value delivered.

The five levels are:

Level 1: The Human-centric Lab
Level 2: The Data-informed Lab
Level 3: The Data-driven Lab
Level 4: The Transforming Lab
Level 5: The Digital-centric Autonomous Lab

Each level is now examined in more detail. The areas of digital skills and enabling IT infrastructure are topics for a future white paper.

Level 1: The Human-centric Lab

R&D labs at Level 1 are dominated by people-centric processes and the scientists who own them. This is nearly every lab that has yet to start utilizing modern digital technologies to enhance their operations. People in the lab take on different roles in order to synthesize, formulate, process, and evaluate materials as part of a broader effort to design specialty products for various applications.

The lab also has "deep experts" who are highly experienced in their domains and who steer the lab in its pursuit of new and improved materials. These experts have amazing intuition about a given design target based on years of building mental connections about what works, what doesn't, and which knobs to turn to reach the desired result.

In specialty chemicals, the experts are relied upon to figure out how to produce a material that meets multiple requirements from a customer with limited time and budget. If one frames the function of the lab as turning chemical structures, formulations, and process parameters into functional materials, the experts reason about "the inverse problem," taking a hypothetical material with desired properties and finding the parameters that the lab can use to create it.

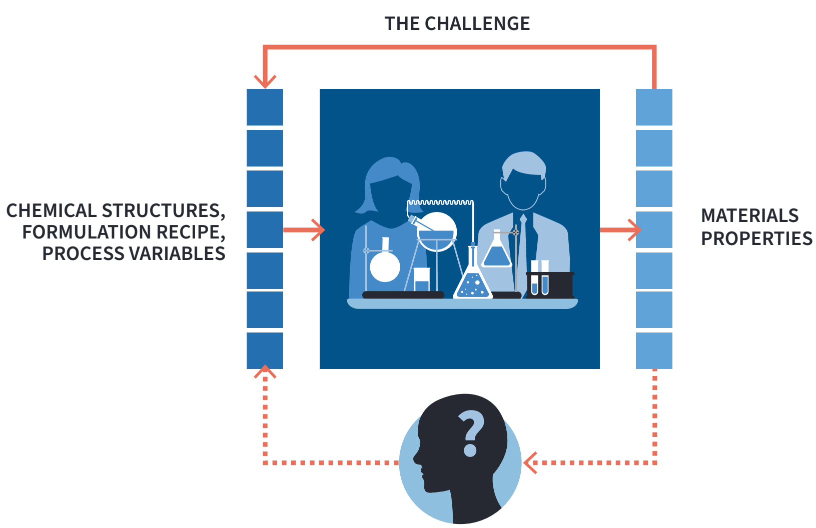


Figure 1: *The role of R&D researchers is to figure out how to take the function of the lab – that of creating materials from chemical structures, formulations, and processes – and invert it to determine the laboratory inputs that will produce a material with the desired properties.*

Lacking well-organized knowledge and data bases, these experts are heavily relied on to get new projects started off on the right foot. However, these experts currently are retiring faster than new ones are being trained to take their place. Even if you are not facing a retirement crisis, there is only so much pattern complexity that can be held in the human mind, and human judgment can be highly biased.

Level 2: The Data-informed Lab

Key data-related advances and significance

- Data is organized and discoverable across the lab
  - Historical data is available for use in visualizations and analysis for improved decision-making
  - Expert efficiency and motivation are increased
  - Perception of data value increases, improving practices and culture

Key technology advances and significance

- Data storage and search
  - All data are saved and usable in decision-making
  - Less time is spent searching for data
  - Laboratory data is saved automatically and accrues value
- Automated data analysis
  - Routine analyses are standardized and performed without human intervention
  - Fewer diversions due to human variability and biases
  - Focus shifts to improving analysis from performing analysis

Key scientific staff elements and significance

- Experts are much more effective with new data access capabilities
- Experts are much more effective in mentoring and knowledge transfer
- Researchers rapidly gain new knowledge and skills, reducing dependence on experts and becoming those themselves

To move from Level 1 to Level 2, the first step is organizing data, new and old, and making it accessible for improved decision-making and results. This is mandatory for all laboratory transformation efforts. There are two aspects of this; one is technical, the other cultural.

All lab processes must put resulting data under centralized management. There are multiple ways to approach this, and not all of them provide sufficient organization and accessibility to aid in decision-making. If the data management solution is not built for scientists, then there's a good chance they won't use it, so there's no point. Data is a corporate asset that has value far beyond the cost of storing and curating it, but only if the data is curated properly.

One must consider how data will be collected and how to minimize mindless data entry and inefficient aggregation of results. This can be facilitated by digital tools that collect data at the source (instruments), allow for *ad hoc* investigation, and automate routine analysis. Once that data is ingested by the system, it must be discoverable so that domain experts can use it to solve problems. This leads to the second aspect: establishing a data culture.

Once that data is ingested by the system, it must be discoverable so that domain experts can use it to solve problems.

Properly managing data in the lab requires cultural change. The perfect data management system will not be worth much if lab researchers and managers do not value data as a corporate asset. At Level 1, data is used primarily to answer some immediate and specific questions, and then effectively lost forever, buried in spreadsheets or presentations, usually on individuals' hard drives. Even if the data files are archived in some way, data not salient to the short-term questions at hand are not likely recorded, and the surrounding context of the data is usually lost.

To move up in levels, the core mission of the lab must change to lead with generating data and knowledge as foundational to creating and evaluating materials. True, materials are still made and tested in the lab. However, the broader R&D decisions about which research direction to pursue next must be derived from the data. This data must be recognized as the most valuable output of the lab and, therefore, treated as such. Learning how to validate ideas and make decisions based on data means centralizing historical data and translating the deep domain knowledge of the experts into a process. Every lab has a few experts

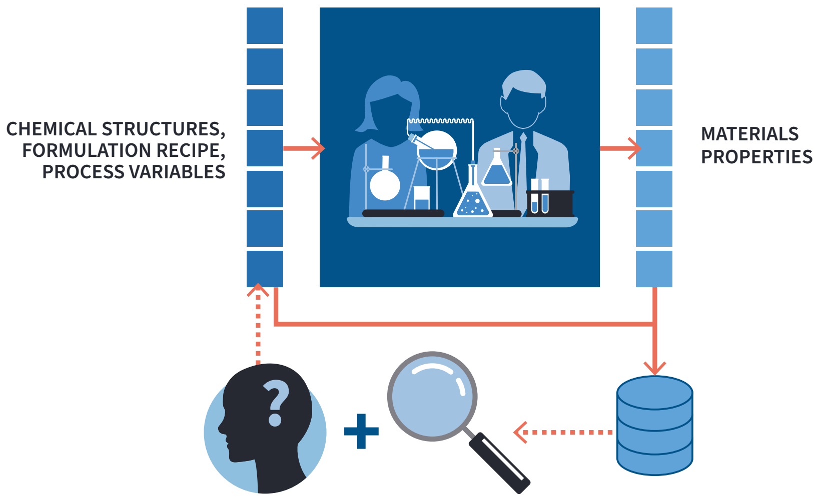


Figure 2: *At Level 2, all data is digitized and centralized, providing search and data visualization capabilities that can be used in decision-making. This focus on the criticality of data to the business requires both technology and cultural change, and is foundational to future advances.*

who have developed intuition after seeing and studying massive volumes of data over long careers. For junior researchers, discoverable data, including all historical data, provides the necessary foundation to both function like an expert and accelerate becoming one.

People will value data more when it can be used immediately to improve their work – a first step in establishing a data culture. The ability to search for relevant data, in particular, is one way to achieve this. Effective data infrastructure will let researchers make queries such as "what materials have we already produced that get us close to these properties?" With this capability, R&D teams can prevent unnecessary duplication of previous efforts in the lab, start projects closer to the target material, and reduce their dependence on experts, thereby making the lab more efficient and resilient.

People will value data more when it can be used immediately to improve their work – a first step in establishing a data culture.

The Level 2 lab with centralized data and search capabilities is more efficient, but the experimental design approach is still very traditional. R&D leaders look at what has been done before, and then put together an experimental plan that systematically varies the parameters that appear to have a significant impact on the target properties. While this process can be roughly informed by manual searching and inspecting historical data, there is nothing fundamentally different from the experimental design strategy used at Level 1. This traditional process can be extremely slow and inefficient, particularly when doing multi-objective optimization over many tunable inputs.

Level 3: The Data-driven Lab

Key data related advances and significance

- Decisions are derived directly from data
  - Adaptive experimental design reduces reliance on experts to guide experiments
  - Design objectives are reached in fewer trials
  - More gets done with the same people in the same lab

Key technology advances and significance

- Adaptive experimental design system
  - Adaptive experimental design recommends best experiments based on *all* managed data
  - Machine-learning-based decision-making is systematic and reasoned
  - System becomes more informed and makes more accurate predictions with more data

Key scientific staff elements and significance

- Experts shift to supervising and tuning algorithms for experimental design to optimize time, cost, and exploration
- Researchers rely less on experts for guidance, freeing experts for higher level tasks
- Junior researchers deliver value with less training and experience

A significant advance is present at Level 3: enhancing decision-making through Machine Learning, utilizing the now-improved and managed data. Most specialty chemical products have numerous degrees of freedom in both composition space and processing conditions, and require optimization to meet several performance and cost objectives. This leads to very slow product development that consumes valuable raw materials, equipment time, power, and worker hours. Specialty chemical suppliers that can conduct this complex optimization process faster and more efficiently can develop superior products.

At Level 3, the digital tool required to excel at complex optimization problems is called adaptive experimental design. The technique is described at a high level in [this case study](). An adaptive experimental design system gives a new superpower to the R&D team. With it, they can leverage the historical lab data to know which direction is most likely to get them closer to the material that the customer desires. In contrast to traditional experimental design methods, adaptive experimental design recommends which experiment to perform next based on a machine-learned understanding of historical data that includes prediction uncertainty. This recommendation is updated after the completion of each experiment so that the best decision is made each time using all of the latest data.

An adaptive experimental design system gives a new superpower to the R&D team.

With adaptive experimental design, R&D leaders can control the amount of exploration in the development process, thereby investing resources today that will improve future product optimization tasks. Laboratory managers also can inject business factors into the system so that R&D decisions are informed by things such as cost, deadlines, and risk levels. With this information, the system can advise what materials parameter values to try, as well as when the likelihood of reaching the design target is low enough that it's no longer worth continuing the project.

It can tell you when to pivot to a new project and even which customer requests you should reject before even starting. Managers also can use it to help efficiently

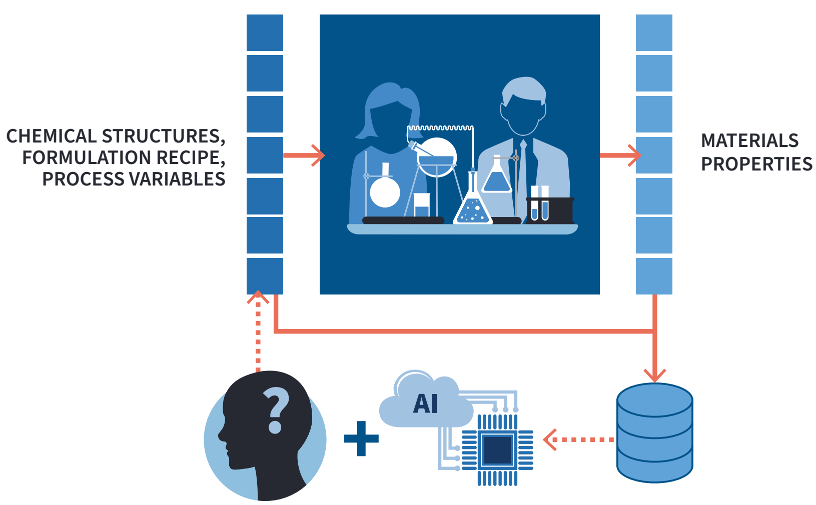


Figure 3: *At Level 3, an adaptive experimental design system is used to help make the best decisions based on all available data, combining what is technically possible with business constraints. This capability initiates a culture shift where R&D leaders learn to trust their data and become comfortable with algorithmic decision-making.*

allocate laboratory workers and other expensive lab resources because the system can learn from previous projects and compare various optimization strategies to be applied to new projects.

Importantly, domain expertise is critical to implementing a successful adaptive experimental design system. With the relatively small data volumes available in most labs, the system is only as good as the quality and choice of data on which it is trained. Domain knowledge is encoded into the system by choosing how to pre-process the data to extract meaningful information (feature engineering) and setting appropriate constraints on the data-driven models to comply with physical restrictions and known phenomena.

Business input also is needed to select key performance indicators and set appropriate rules that help the system make informed recommendations that capture various business factors.

At this point, the materials design process is digitalized. The lab is much more efficient, owing to an ever-growing collection of valuable centralized data and a constantly evolving system supporting R&D decision-making. However, nothing has been fundamentally transformed about the process. It is still a human-centric laboratory, and the R&D process flow is largely untouched, which puts a ceiling on the business value.

Level 4: The Transforming Lab

Key data-related advances and significance

- More data and more types of data are being generated by the lab
    - Squeezes the most out of the lab and people without a complete redesign
    - Achieves results faster and better due to more information
    - Focuses changes on improving processes vs. running processes

Key technology advances and significance

- New instrumentation and measurements, tailored to feed knowledge
    - Faster, cheaper, and more informative data for a more informed adaptive experimental design system and better materials
- Digital twin
    - High-speed, low-fidelity data for materials screening
    - Explores materials solution space before iterating in the lab
    - Provides hidden descriptors that cannot be measured
- Experimental automation prototypes
    - Faster and higher-quality data for the machine-learning-assisted decision-making

Key scientific staff elements and significance

- Experts identify new and better measurements, reducing reliance on intuition and bringing experimental results closer to the experimental design
- Experts and researchers identify process improvements that provide more and better data to accelerate laboratory operations
- Researchers execute on experiment designs created by the adaptive experimental design system; the experts are involved by exception

Level 4 is the stage where laboratory transformation starts, with a vision for a new digital-centric autonomous lab that leverages current and future technologies fully. This vision is necessary to achieve transformational gains. At Level 3, the lab is significantly faster and more performant than it was at the caterpillar of Level 1. Relatively speaking, however, it is still a caterpillar.

Level 4 is the stage where laboratory transformation starts.

It is time to create the vision of a butterfly.

The adaptive experimental design system has given the lab the ability to learn from history, turning data into knowledge. With this system making use of the data that everyone knows, researchers can start thinking about new data sources. This system will eventually run the lab. Before it can, it needs more high-quality data and more insightful data that is generated faster.

One place to start is to capture the intuition of the experts. What is that intuition? It is special but surprisingly mundane. A good expert serves as another measurement device for the lab. For example, an expert may know that a certain unmeasured smell or color means the temperature got out

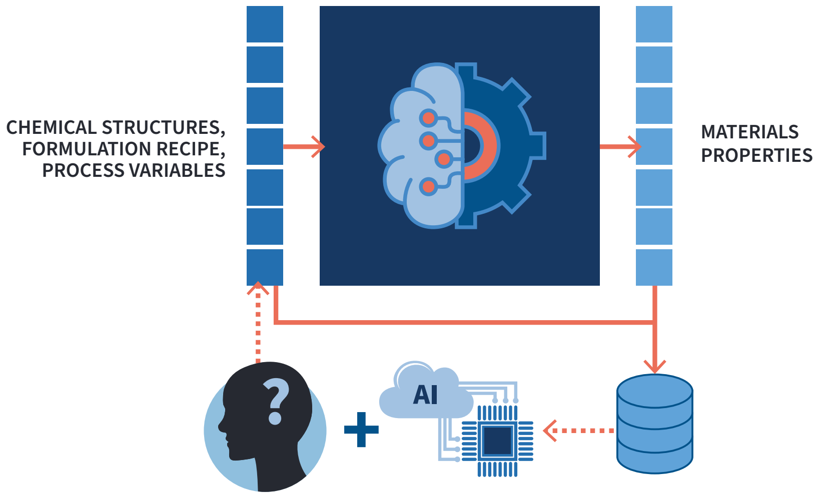


Figure 4: *At Level 4, the ideal autonomous lab is envisioned and explored. Experts develop new descriptors that capture their intuition, new measurements are introduced, and portions of the laboratory process may be automated. The digital twin of the lab also can be created, augmented by physics-based simulation, and used to provide a wider view of the design space. The key cultural challenge at this level is getting comfortable with re-thinking decades-old R&D workflows and being willing to prototype new solutions to sample fabrication and characterization tasks.*

of spec during formulation. Whatever it is, the expert can detect it. Capturing this may mean that the lab needs to track temperature more closely during a chemical reaction or mixing process.

It is time for the experts and the best analytical R&D minds to come together to identify the new, better, or faster measurements that will help accelerate and redesign the R&D workflow. This also might entail a new proxy measurement performed early in the workflow to screen out 90% of the bad candidates before the slower, more-costly measurements are performed.

To move toward the ideal digital-centric autonomous lab, a lab also must start identifying process bottlenecks that can be alleviated using robotic automation. Whether it's using high-throughput sample creation, sample characterization, or both, it's time to start envisioning and building modular, automated laboratory equipment that is digital-centric in its design. These tools must be able to receive commands from and send data back to a central control system, and they must be reconfigurable and interoperable so that they can be repurposed as the lab workflow evolves.

Physics-based simulation is an additional source of information available to R&D labs. So far, the adaptive experimental design system has been built around experimental data, but it works just as well wrapped around simulation data. Various properties and processes in the lab can be probed virtually using physics-based simulation models to form a digital twin.

... a lab also must start identifying process bottlenecks that can be alleviated using robotic automation.

The digital twin can be a powerful tool for scouting out areas that may be very expensive to explore experimentally. In this paradigm, the digital twin can be thought of as a lower-fidelity map of the design space. One can go to the map to see what they might find in an area that has yet to be thoroughly explored experimentally. Simulations can also reveal hidden fundamental materials properties and process descriptors that are not easily measurable in the lab.

A key attribute of the digital twin is that it can run non-stop at a much lower cost per data point than many experiments. Even though the digital twin is imperfect, the adaptive experimental design system can use it to help make decisions about the real lab. The system can learn the correlations between the digital twin data and the experimental data, and can determine at any point in time which fidelity of simulation or experiment to run. In this way, limited computational, lab equipment, and human resources all can be allocated efficiently to extract maximum knowledge given the data gathered so far, the design targets, and the project constraints.

A lab may remain and improve continuously at Level 4 for a considerable amount of time, perhaps several years. All the technological and cultural factors needed for an autonomous lab are identified or in place, with the next step being end-to-end automation of real-world lab processes. A central control system guided by adaptive experimental design may be driving major portions of the lab with little human input, where some parts of the process may not have been automated due to priority, cost, or trust in the system's capabilities. Automating the entire lab will take more work, but what was impossible at lower levels is now within reach.

During this period, the vision for the future autonomous lab is created and with prototyping carried out, completely removing humans from a new set of innovative workflows.

Automating the entire lab will take more work, but what was impossible at lower levels is now within reach.

With some digital skills training, agency in the technological innovation, and cultural changes taking place in the lab, the experts that once directed experimentation are now leading the transformation and how this transformation is affecting the customers of the lab.

Level 5: The Digital-centric Autonomous Lab

Key data-related advances and significance

- Lab redesigned and reimplemented around generating data
  - Data generation is no longer a human-driven process
  - Data is cheaper and generated faster than ever before

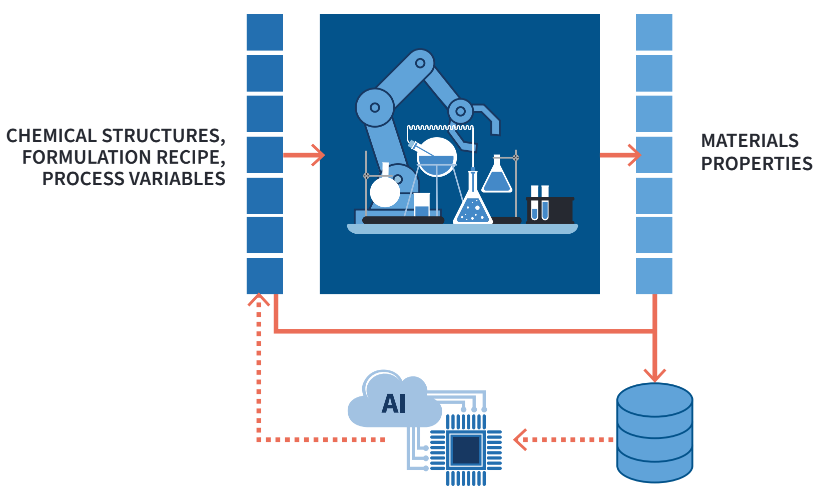


Figure 5: *The learning, technology, and experimentation of the previous levels have brought the lab within reach of being fully automated. At Level 5, labs integrate these disparate capabilities into autonomous systems that can evolve as business needs change. These systems confidently produce new materials that meet customer specifications orders of magnitude faster than before and enable broader business transformation.*

- Data is tailored for data-driven decision-making

Key technology advances and significance

- Complete process automation
  - Culmination of learning, prototyping, and ideation at Level 4

Key scientific staff elements and significance

- Experts are guiding other transformations and thinking ahead to the next generation of the lab
- Researchers become feature explorers in the automated lab and still guide non-automated work to gather rarely needed but uniquely valuable information
- Technicians shift to testing, maintaining, repairing, and upgrading lab equipment

Level 5 completes the transformation started at Level 4 into a digital-centric autonomous lab. Lab automation requires new or reconfigured instrumentation that is fit for how the digital-centric lab needs to operate. This new lab equipment may be purpose-built and relatively inexpensive (see The 1000x blog post) or an assemblage of off-the-shelf equipment from traditional vendors. In either case, it will be foundational to the digital-centric modern lab.

Use of the adaptive experimental design system to explore the design space, both experimentally and with digital twins, has informed on data necessary to produce high-value products for targeted markets. By tweaking the system and models, trust and confidence are gained in its capabilities.

One likely outcome is discovering that not all of the earlier measurements are necessary to arrive at good results, at least not most of the time. These earlier measurements will still have an important role, but often only after 95+% of the unsuccessful materials and chemical candidates have been eliminated through the automated processes. These rarely used techniques may well not be candidates for automation. However, the system can still identify special circumstances and recommend humans to perform these more exotic measurements.

It is important to examine how the hours gained through efficiency will be spent by the experts, junior researchers, and technicians. The lab has been transformed dramatically and now automatically generates the vast amount of data required to serve both its historical business function and create new possibilities. The automated processes and enlightened people are ready to deliver new areas of innovation for the business.

At Level 5, continuous process improvement must become a core cultural element to make the autonomous lab and the business as a whole robust and resilient. On the data side, data-driven models are highly dependent on feature selection and feature engineering tasks. As the overall process accelerates, there always will be bottlenecks where the adaptive experimental design system simply struggles to predict an important property in an economical way.

Lab scientists must transition to feature explorers and creators, where deep domain expertise is leveraged to continually create new measurements and data analysis methods that increase overall predictive power. There also is a need for skilled engineers to tackle process improvement tasks – for example, increasing measurement accuracy/reliability and decreasing the time and cost of existing measurements and simulations, making data generation both more efficient and improved for the targeted experiment. For lab technicians, equipment calibration, maintenance, repair, and testing will all still need to be performed regularly.

> Use of the adaptive experimental design system to explore the design space has informed on data necessary to produce high-value products for targeted markets.

There also will be opportunities for experts to provide leadership to transform other parts of the company, including beyond the R&D lab. They may gain numerous non-technical skills through this lab transformation, unveiling new career opportunities.

The role of R&D management also will undergo significant changes. The lab will now be delivering not-seen-before results to the larger organization. Management will need to collaboratively evaluate new business potential – for example, the new ability to meet a client's custom chemical request in weeks or days vs. months may open up new market opportunities and require new ways of interacting with customers.

Final Thoughts

New digital technologies are frequently introduced into labs with associated incremental performance improvements and value. These technologies are often selected based on a clear path to value using existing workflows, such that organization and cultural changes are minimal. However, without changing workflows or culture, digitalized processes remain limited by their current forms.

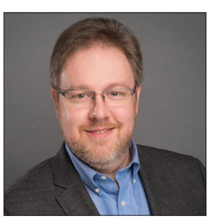
To have a transformative impact, labs must reinvent workflows and adopt a strong data culture. Researchers must acquire new skills and be empowered to bring digital innovation into the lab. Digital technologies must be selected that can rapidly evolve in step with the lab. A lab R&D system that is too rigid, inefficient, or adopted as a quick fix, must be avoided, as it will be incapable of broader transformation and unable to adapt as business needs change.

When the lab arrives at a point where scientists can dial-in desired material or chemical properties, and samples with those properties are produced quickly and automatically, there has been a true transformation. It is now possible to develop highly customized products for each customer, bring speciality services into new markets, and stave off commoditization. From there, the business must decide how to leverage this new capability. The challenge flips from a technical one of creating samples, to a business one of scaling production capacity, creating new customer-focused digital sales tools, expanding into new markets, and generating increased revenue - a good problem to have.

Key to advancing to a Digital-centric Autonomous Lab is that technological and cultural changes progress concurrently. The technological initiatives generate value, while the cultural and organizational ones accelerate it, increasing its potential beyond incremental steps, and ensuring a foundation for future progress. Once a given level is mastered, the lab is positioned to move to the next. At the final level, entirely new possibilities can be explored and a new future envisioned in line with broader digital business transformation goals.

Thinking more broadly, it may become possible to vertically expand the business beyond being a specialty materials or chemicals supplier, to one of end-use products that are enabled by self-generated, innovative materials. There may also be new value-added digital products and services to offer along with the original product, now enabled by the autonomous lab. We'll save that as an area for future exploration.

About the Authors



Chris Farrow is VP, Materials Science Solutions, and holds a a Ph.D. in physics from Michigan State University and degrees in physics and mathematics from the University of Nebraska.

Michael Heiber is an Application Engineer, and holds a Ph.D. in polymer science from The University of Akron and a B.S. in materials science and engineering from the University of Illinois at Urbana-Champaign.

NORTH AMERICA
Austin, Texas
Headquarters
200 W Cesar Chavez, Suite 202
Austin, Texas 78701
United States
+1-512-536-1057

Houston, Texas
1502 Sawyer Street, Suite 232
Houston, Texas 77007
United States
+1-512-536-1057

EUROPE
Cambridge, United Kingdom
Broers Building, 3rd Floor
21 JJ Thomson Avenue
Cambridge CB3 0FA
United Kingdom
+44-1223-321679

Zürich, Switzerland
Kalanderplatz 1
8045 Zürich
Switzerland
+1-512-536-1057

ASIA
Tokyo, Japan
3/F Shiodome Building
1-2-20 Kaigan, Minato-ku
Tokyo 105-0022 Japan
+81-03-6721-8787

info@enthought.com
enthought.com

ENTHOUGHT