

The Coming Revolution in Materials Informatics and Artificial Intelligence

Authors: Greg Mulholland, CEO, and Douglas Ramsey, Vice President

Citrine Informatics, 1741 Broadway, Redwood City, CA USA

CITRINE INFORMATICS: THE DATA-DRIVEN PLATFORM FOR ADVANCED MATERIALS

Artificial Intelligence (AI) and Machine Learning (ML) is the next foundational technology revolution that will reshape how we work, live, and organize our societies. Citrine is helping architect this world of the future by creating standards, protocols, and AI engines for materials discovery and the manufacturing of those materials.

Citrine's primary goal is to help our customers reduce their materials discovery, costs, and development schedules by 50%. Citrine uses Artificial Intelligence (AI) engines to achieve this result or better for several government and commercial customers. Our AI tools sort through our databases to radically narrow our customer's research areas to make sure that they don't waste time and money pursuing less promising areas. Citrine does not seek to own our customer's data or discoveries. Rather Citrine offers a robust Materials Informatics (MI) platform and AI analytical tools that help our customer make new breakthroughs in half the time at have the costs. Our customer's data and discoveries always remain their intellectual property. Citrine is successful when our customers are successful and create new and novel materials, products, and solutions.

To support our platform, Citrine is building the world's largest materials data platform. Today, our system has over 18,000,000 material-property pairs. There are numerous public and private materials data repositories within the Citrine platform. Citrine provides customer specific research domains for our customers which allows them to house their proprietary data and combine their private databases with Citrine's larger materials data platform.

Citrine's platform encompasses data for a wide range of public and private repositories and pulls them together into a standard and readily searchable format for researchers,

manufacturers, and designers. The database is continuously growing with support from our academic, government, and corporate partners around the world. Today, any researcher, in the US or overseas, can use this database to find information about their field and materials that have been studied. Citrine allows our customers to rapidly sort through millions of data points to make sure they are focusing on the right challenges and improving their time to discovery. At Citrine we are passionate about driving improvements and development of AI tools to improve industrial efficiency, output, and growth for our users.

We believe that our suite of tools are fundamental enablers that will drive the next industrial revolution often referred to as 'Industry 4.0' or 'Second Machine Age'. Citrine is uniquely focused on challenges related to materials discovery, design, and manufacturing. Good products start with great materials. Citrines brings together material science, design, manufacturing know-how, and computation to create new insights and new value for our customers. Ultimately, Citrine believes in the power of our products to 'do good' for society by creating new, more environmentally friendly products and new industrial disciplines that will promote new job creation across both new and traditional industries.

"Our data management practice is for a scientist to open Microsoft Word, edit a file, save it to PDF, and email it to a webmaster."

—Director, Fortune 100 Materials Company

ORGANIZE THE WORLD'S MATERIALS DATA AND BREAK DOWN SILOS

The days of having a scientist go to a shelf to pull a book or scour archives for a PDF are, surprisingly, still not in the past at most companies. Citrine obviates this need by ingesting data. Our automated tools reach into documents and pull out relevant quantitative

data. We preserve the original document but make data that had been locked away in narrative text, charts, and figures available to both humans and computers alike in the most analyzable format possible.

Citrine does not stop at text, though. In talking to over 100 companies, we discovered that the norm in the industry is a scattered array of digital data formats from measurement tools, manufacturing and synthesis systems, and spreadsheets. Citrine's systems ingest the files, and makes each development step more computable than ever before: enabling the comparison of results across organizations. Manufacturing companies, due to their size and nature, become incredibly siloed over time. Groups working on related problems often are unable to share information with one another. This issue is exacerbated by the fact that few companies have data systems that facilitate natural sharing among these groups.

CREATE A CORPORATE ASSET FROM DATA

More importantly, though, is that materials development is iterative. If a product developer spends two years working on a project trying to create new material and tests 10,000 candidates, and one succeeds, that one success is celebrated.

Citrine's platform makes the whole process valuable. By capturing the exhaust of the development process, we are able to make those 9,999 "failed" experiments into a valuable asset. Citrine's platform puts this data into motion, and enables much more advanced analyses. Rather than seeing a company's development history as a series of failed tests and dead-end development pathways punctuated by an occasional success, Citrine enables the development process to illuminate the world of materials ever more quickly to make decisions ever faster.

COMBAT CONFIRMATION BIAS AND OPTIMIZE PRODUCT DEVELOPMENT

Today's product development pathways are driven largely by intuition. The career history of a scientist, coupled with a confirmation bias based on what has worked most recently, informs next steps in an outsized way.

Surely, scientific intuition is a component, but Citrine brings another dimension: massive datasets analyzed in ways never before possible.

"I can look at a [bill of materials] for a recipe and tell you who wrote it...people have their favorite tricks."

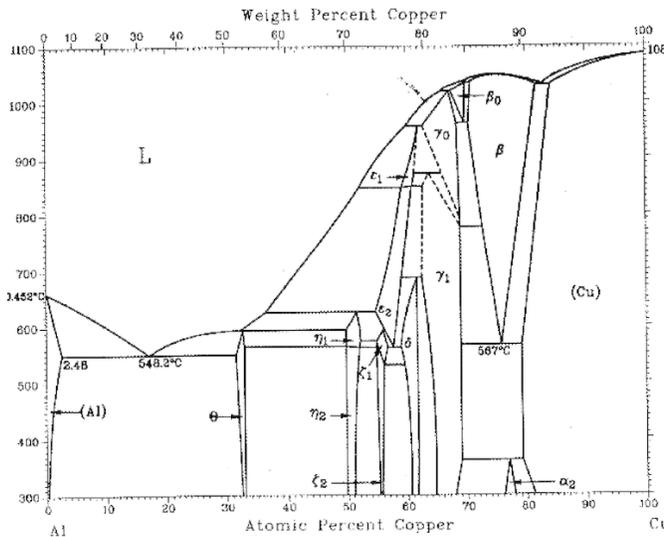
—Director, Fortune 500 Materials Company

Citrine allows product developers to use a data driven method determine what gives us the best possibility of achieving the target characteristics. These systems can simultaneously recognize the trade-offs between various properties and identify complicated trends in the data so that humans can do what they do well: using their intuition and understanding of materials selection within a narrow scope and connect those materials performance characteristics to specific business needs.

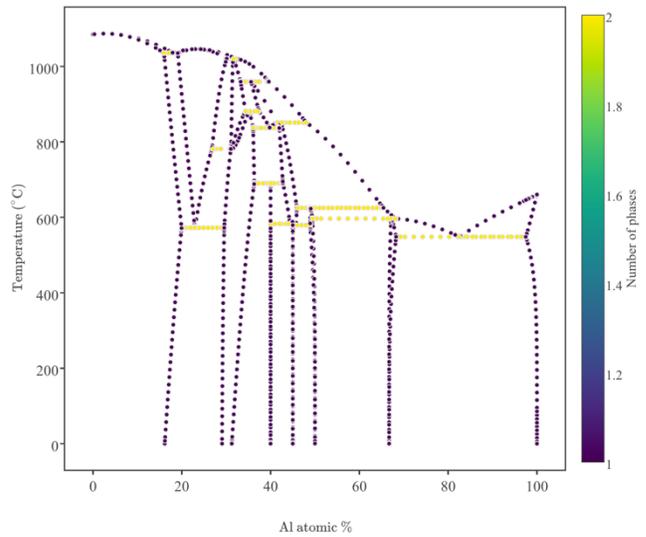
Most importantly, every time the product developer iterates, Citrine's platform learns and improves. As such, the platform learns from all historical data as well as each incremental development step, constantly making the development process more and more efficient. Citrine is not a one-time investment in software; Citrine enables a new paradigm of product and materials development.

CITRINE'S ADVANTAGES

Citrine's advantage is the combination of three things: the world's largest repository of materials and chemicals data; cutting-edge, secure infrastructure to support that data; and unique machine learning systems to accelerate the discovery of new materials and chemicals.



Traditional analog phase diagram image.



Citration digital Al-Cu phase diagram.

1. World's Largest Materials Repository

Citrine hosts a community platform for data storage and analysis called "Citration". This system is provided free of charge to universities and national labs. In return, the Citration platform is allowed to learn from that data to improve its algorithms. This data set grows continuously as new work is published to the platform.

Citrine also has data sharing partnerships in place with some of the most well-known authoritative materials publishers including: US National Institute of Standards and Technology, American Society of Metals, MMPDS, and many others.

2. Intelligent Materials Data Infrastructure

All analysis methods require input data that is in a machine-readable, structured format. Materials data, in particular, can be exceedingly complex, so defining data formats to store any and all materials-related information is a daunting task. Citrine has developed a hierarchical data structure used for storing materials data called the physical information file (PIF). The PIF is a flexible schema for storing the structure, processing history, and properties of materials, devices, and physical systems. We have built a powerful data control system on this format that enables the rapid searching, analysis, and visualization of that data.

3. Machine Learning for Materials

Citrine uses state-of-the-art machine learning systems, optimized to take advantage of trends in chemistry and physics to combine computational simulation, traditional analysis, and experimental data into a powerful prediction engine. This engine can be used to predict the performance of known materials under new conditions, optimize new materials for known applications, or optimize the development process to reduce the risk and total number of iterations to achieve a product goal.

BUSINESS MODEL

Citrine Informatics is a Software-as-a-Service (SaaS) business and deploys cloud-based data infrastructure, analysis tools, and machine learning systems. Subscriptions are changed on an annual basis. Citrine makes no claim on data or materials IP that our customer discovery via Citration. Rather, Citrine is committed to ensuring that our customer retain all rights for new breakthroughs to enable them to accelerate their time-to-market for new materials and applications.

"I got six additional papers out of using the Citrine platform. This collaboration was the basis of my NSF Career Award."

—Professor Taylor Sparks Univ. of Utah

MARKET

Citrine sells to the Global 1000 companies involved in the development and manufacture of advanced materials. In addition, we service Materials-Enabled Product companies (MEPCs) like large automakers, electronics manufacturers, aerospace companies, and many others. The most innovative companies in each of these industries spend up to 10% of revenue

on their research and development efforts, yielding large opportunities for improved efficiency. In addition, the value of a new product delivered to market a quarter early is often worth hundreds of millions of dollars in revenue and even more in market capture. Given that Citrine has delivered a reduction of 50-75% in development time within these very large industries, the demonstrated value and market is extremely large.