

White paper

**AMPLIFY THE IMPACT OF YOUR  
DATA SCIENCE TEAM**



# Amplify the impact of your data science team

## INTRODUCTION

You know that Materials Informatics can add value to your company... You have a smart team of data scientists, data analysts, and simulation experts... There are plenty of open source data management and machine learning tools available online... It's tempting to knit these tools together and build an entire MI infrastructure from scratch... In fact, you may have run a successful in-house pilot project with open-source or custom-built tools.

### This paper covers:

1. Why it's challenging to recoup your investment when building your own MI infrastructure.
2. How the Citrine Platform amplifies the impact of your existing data science and MI teams.
3. Recommendations for onboarding and implementing a 3rd party MI platform with your data science and/or MI teams.



**A recent market report by Algorithmia<sup>1</sup> found that “38% of organizations spend more than 50% of their data scientists’ time on deployment—and that only gets worse with scale.”**

**The same report, which surveys company’s machine learning initiatives across different industries, also found that companies saved around 20% on infrastructure costs annually by buying third party systems and speeding up the time needed to deploy a model.**

<sup>1</sup>[2021 enterprise trends in machine learning](#)

## 1. DIY?

Your team is smart and enjoys a challenge, so there's no question that they can leverage in-house or open source tools to solve a project-level R&D problem with MI. However, if your goal is a digital transformation of the way you discover, develop, and deploy new materials across your business, here's 3 things to think about.

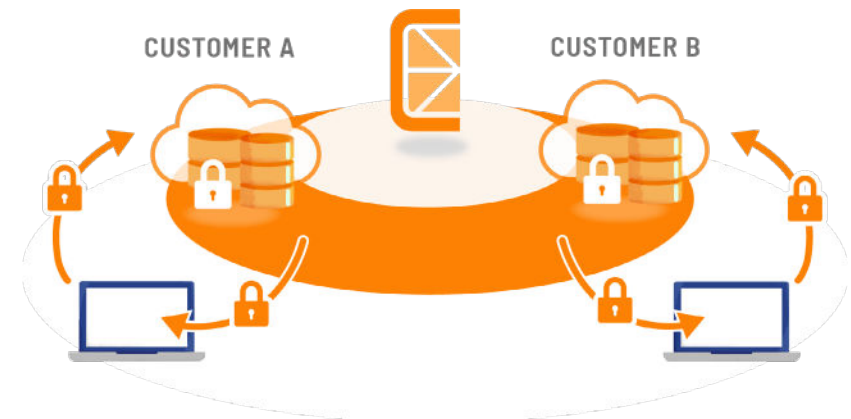
1. Can your in-house data science or MI team build a scalable, safe, trusted system that engineers and scientists reliably use on a daily basis?
2. How long will it take for your team to put it in place? Is it worth waiting for? Or do you want to get value from MI now?
3. If you use generic AI and data management tools, how will you make them work with complex materials and chemicals data?

### Your platform must be secure

Many AI algorithms are open source, but your data is unique. This hard-earned IP is what will power your MI program and deliver sustainable competitive advantage. It should be protected. Citrine has put the work in to make sure that the Citrine Platform follows security best practices. [Citrine is ISO27001 certified](#) and works proactively to identify and prevent security threats. Data on the platform is encrypted and segregated and an authorization system enables you to control access to data across different teams and business units.



### Segregation and Encryption



How will you secure your data?

Will that still work as other business units want to use the system?

## 1. DIY? (cont.)

### Your platform must be scalable

Many data management and machine learning tools are available open source, and for a pilot or proof-of-concept project the overhead of stringing these together could make sense. But does it still make sense to do this on projects 2, 3, 4...

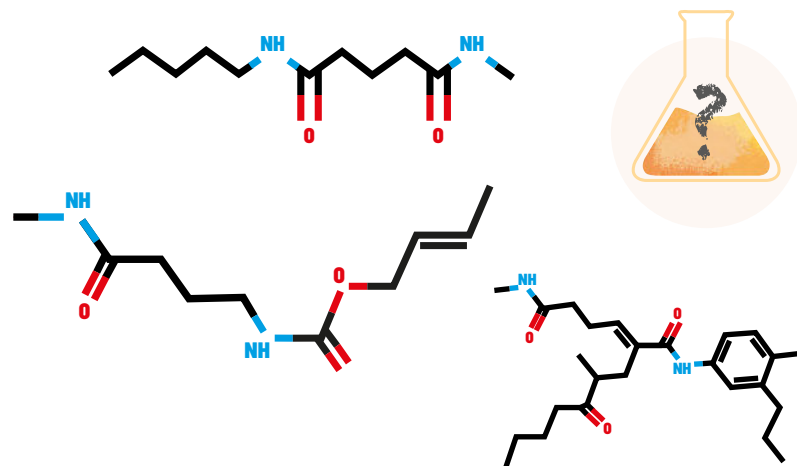
Without being able to reuse digital assets, each additional project has to start from scratch. The code quickly becomes a morass, that the main architect can understand, but it is impenetrable to new starters on the team.

### Don't add another layer of complexity

AI workflows are complex and expressing materials problems through a general-purpose AI solution adds yet another translation layer. The Citrine Platform is a materials-aware AI solution, and offers native support for materials data types and formulations problems. That means that with Citrine, you can get up and running developing new or optimizing existing formulations, without worrying about infrastructure, translation layers, or repeatability.



Will a new team member be able to understand what has been done previously? How long will it take them to get up to speed?



How will you featurize chemical formulas?

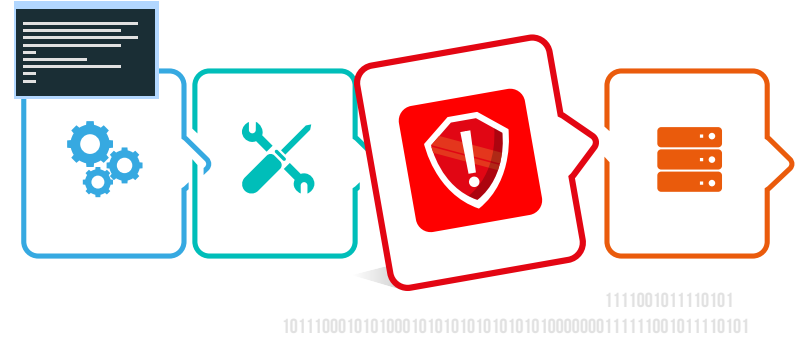
## 1. DIY? (cont.)

### When you build a system, you must maintain it

Our team is constantly testing and improving our platform with faster infrastructure, cutting-edge algorithms, and improved UI. They ensure uptime and site reliability and update the platform as integrations, databases, and cloud infrastructure continue to evolve. While this isn't headline grabbing work, it is essential for a software system that can continue to be used over the medium and long term.

### No IT system exists alone

Integrating with third-party software, such as testing equipment and simulation packages requires rational API design, data pipelines, and data standards. These are the expertise of Citrine. An ad-hoc API would not be stable enough for business process integration.



Is maintaining a software system something your team would enjoy doing? Can you afford an ongoing team of software engineers?



## 2. AMPLIFYING THE IMPACT OF YOUR TEAM

If you are not building or maintaining your own software – what can your team do instead?

### You can focus on accelerating product development

With time freed up from infrastructure, security, and code maintenance activities, your team can focus on more materials-related problems.

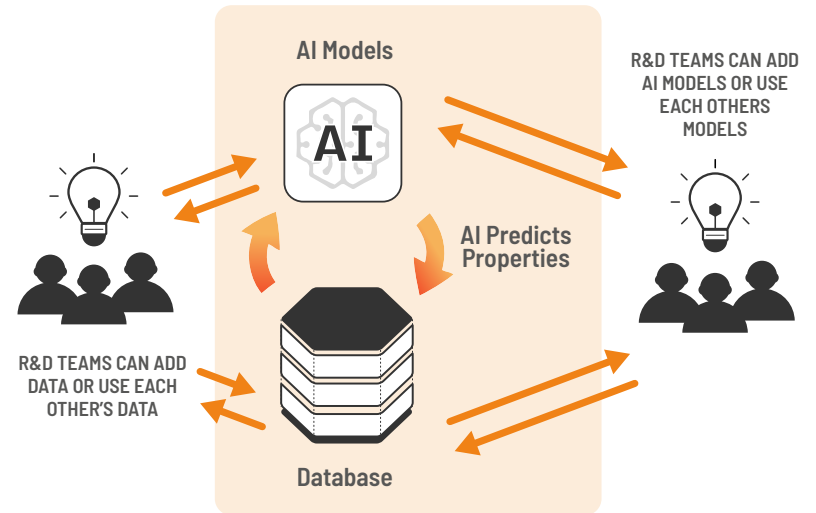
### You don't lose what you have already developed

You can use your data, and data from any simulations you have developed in our platform. You can also make an impact across your organization by making your work available to others.

### Reusable assets and automation will save your team time

Automated data pipelines and workflows will save your team time and make sure that your team is always working with the latest data. Datasets, design spaces, and AI-models are all purposefully designed to be reusable, not just by your team, but by authorized members of other business units.

### No need to reinvent the wheel



## 2. AMPLIFYING THE IMPACT OF YOUR TEAM (cont.)

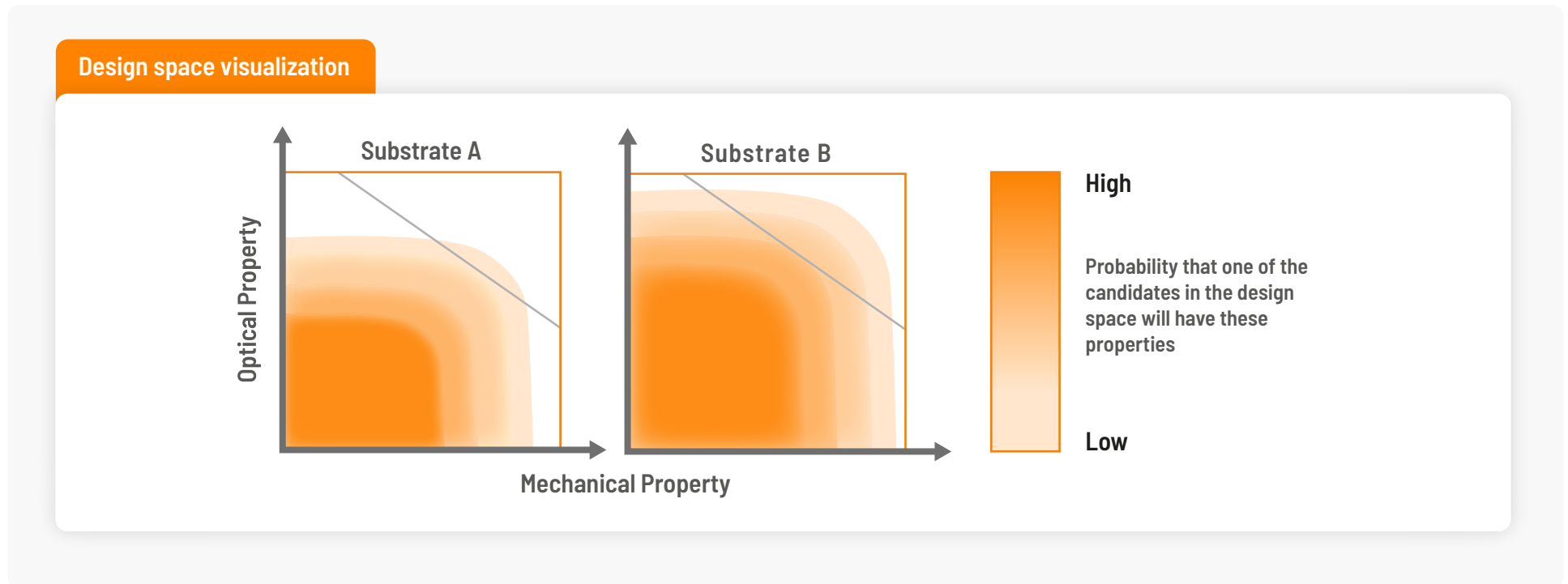
### Your team can focus on becoming experts

They will:

- Learn how to use tools on the platform to create design spaces; partnering with materials domain experts to understand relevant constraints and input parameters.
- Capitalize on the domain knowledge of the wider team to improve model accuracy by learning how to create hierarchal machine learning models incorporating analytical relationships.
- Become familiar with different search strategies and how best to use them on different R&D projects.

### You can objectively guide research directions

Using design space visualization, your team will be able to quantify the probability of finding a material to hit targets in different scenarios. You will be able to ensure that business leaders have the objective data they need to make good decisions on whether a research project should start, stop, or continue.



### 3. HOW TO GET STARTED

#### Working with an experienced partner

Citrine has been working with companies on materials informatics projects for 7 years and now has 50+ projects under their belt. Training is incorporated through hands-on collaboration and educational opportunities during onboarding and pilot project(s).

By initially identifying a high-value, AI-ready project, your team will quickly learn the ropes of the new system, prove that the system produces return on investment, and together with Citrine identify the best data structure for your organization.

On the initial project, Citrine works with your team to:

- Ingest data into the Platform
- Build data pipelines
- Train key team members on platform usage and MI methodologies

Citrine will work with your team to progress platform autonomy in a manner that best suits your organization's needs. This includes understanding the capabilities of your R&D team members and assessing their comfort level with platform tools. Developing the internal capabilities of our customers is our top priority, to ensure self-sufficient value creation from Citrine's technology.

#### Route to autonomy

- 1 IDENTIFY HIGH-VALUE, AI-READY PROJECT
- 2 PREPARE DATA
- 3 CREATE, REFINE AND VALIDATE FIRST MODELS

Develop internal capabilities in using the platform

Onboard further projects and business units



## SUMMARY

Creating and maintaining scalable MI infrastructure is challenging. Citrine has built and continues to evolve the leading Materials Informatics System. By partnering with Citrine your MI, data science, or data analysis team can focus on high-value activities like using data to understand customer requirements, making complex business trade-offs, and capturing their scientific domain knowledge in digital tools.

By taking a DIY route, you may have success on the first couple of projects, but your team will eventually focus much more of its time on software maintenance, version control, deployment, security, and infrastructure management rather than value-add R&D activities.

**Contact Citrine to learn more about what the Citrine Platform can do and how we partner with customer teams to help them transform their digital R&D efforts.**



Arrange an online meeting



Listen to our podcasts



Download more case studies



Subscribe to our newsletter

Citrine Informatics Inc.  
2629 Broadway St  
Redwood City, CA 94063

[citrine.io](https://citrine.io)  
[info@citrine.io](mailto:info@citrine.io)

© 2021 Citrine Informatics Inc. All Rights Reserved.