The data-informed institution

How education is using data to accelerate digital transformation

By Mark Schwartz, Enterprise Strategist, AWS

Today's education environment

Like other industries, higher education is in flux, driven by rapid changes in demographics, technology, student needs and preferences, budget pressures … and a pandemic. And as with many industries across the public and private sectors, responding to these pressures will require adaptation through better use of technology and data.

In the United States, the decline in enrollment that began before the pandemic has worsened, with enrollments falling 3.2% from 2020 to 2021, and a startling 14.1% for community colleges[1]. This, in turn, has led to staff reductions of 650,000 jobs between March 2020 and February 2021—a 13% workforce reduction[2]. Attrition is likely to continue, particularly among technologists, with 40% of IT staff saying they are somewhat or very likely to seek other employment opportunities in the next twelve months—just at a moment when technology is becoming critical to delivering educational outcomes[3]. In the UK, thirteen universities were at risk of bankruptcy in 2020, according to the Institute for Fiscal Studies. In Australia, universities may have lost up to $4.8 billion in revenue by the end of that year, with international student visa holders in Australia declining by 54%[4]. Higher education institutions in these countries will be challenged to reverse these declines with fewer resources.

Responding to both mission and financial imperatives, higher education faces pressures to improve student outcomes and increase retention. Colleges and universities are making a concerted effort to increase diversity, provide access to education to a broader population, and address inequities worsened by the pandemic. Given all these trends, higher education providers find that they need to think differently about supporting students through their educational journeys, personalizing the learning experience, assessing risk, offering more flexibility in learning modality, improving the campus experience and providing online support, and attending to mental health. They also need to respond to the budget pressures of uncertain enrollments and rising costs by maximizing their financial resources through diversifying revenue streams, increasing fundraising, and striving for operational excellence—delivering student outcomes in the most streamlined, effective manner. And, of course, they need to make learning available through online and hybrid channels, as well as in person.

1. https://www.highereddive.com/news/undergraduate-enrollment-drops-32-this-fall-deepening-last-years-losses/608854/
2. https://www.insidehighered.com/blogs/leadership-higher-education/650000-colleagues-have-lost-their-jobs
3. https://er.educause.edu/articles/2021/9/educause-and-cupa-hr-quickpoll-results-the-misalignment-of-preferences-and-realities-for-remote-work
4. AWS report, Emerging Trends in the New World of Education (2020), https://www.universitiesaustralia.edu.au/media-item/covid-19-to-cost-universities-16-billion-by-2023/

Content elements:

This is all to say that higher education needs to adapt effectively to rapidly changing circumstances. Traditionally, speed and flexibility have been less important in higher education than in other industries, especially when it comes to technology. But the fast pace of change is likely to continue, and slow cycles for technology change and reliance on massive, expensive IT initiatives are no longer feasible. As a result, **78% of college and university presidents now view the pandemic as a catalyst for transforming their institutions, and 34% view it as an opportunity to "reset for growth,"** according to a survey by Inside Higher Ed in 2021[5]. And 52% specifically link these changes to speed, admitting that they need a governance structure that allows for quicker decision-making. The organizations and institutions that adopt and adapt to technology integration the fastest will be the ones who build the future.

At the core of this transformation is data. Organizations that master the fast and adaptable use of data, as we'll see later in this white paper, can use that capability to address many of the challenges of higher education today. But doing so requires more than just implementing technology. It requires new ways of solving problems, governing initiatives, organizing employees, and fostering innovation.

> In this white paper we'll ask the question:
>
> **How can higher education use data to adapt effectively to change, and how can higher education build adaptability and speed to make better use of data? In other words, how can higher education become more agile in its use of data, and how can it use data to become more agile and adaptable?**

5. https://www.insidehighered.com/news/survey/survey-shows-college-presidents-emerging-covid-19-more-confident-their-institutions-can

Content elements:

The mission and business value of data

Data can be used by educational institutions to better retain students, foster diversity, support online and hybrid learning, drive fundraising activities, manage facilities more sustainably, personalize learning experiences, achieve operating efficiencies, and advance student outcomes. But institutions have often found it difficult to use data in these ways because they thought of data solely in the context of operations and compliance; as a result, they locked it away in siloed databases that were excellent for operational processing but less suited to open-ended analysis. Because the mental model was that of operational aspects of student information, class registration, billing, or course scheduling, data was siloed into student information systems (SIS), learning management systems (LMS), and other operational- and compliance-oriented databases. But using data to inform decision-making and drive automated activity requires a different way of thinking about it, and ultimately a different way of organizing it and working with it. Data is no longer about one-off "transactions" like matriculation or course completion but the subject of analysis and interpretation. Data has value beyond its operational use.

How can we measure this value, and how can we maximize it?

Each piece of data can be used in any number of analyses that will drive mission or business results. We can think of data, then, as an asset whose value lies in making possible the mission results that follow from those analyses. A few examples help illustrate the possibilities:

> **A leading Australian university** worked with Amazon Web Services (AWS) partner **Versent** to plan targeted intervention programs for student support and retention. They discovered that 16% of the student population was at risk of dropping out. The intervention programs they designed are expected to improve student satisfaction by 14% and have the potential of avoiding revenue losses of up to $189 million.

> **Oklahoma State University in Oklahoma City (OSU-OKC)**, a four-year institution in the United States, used data to benchmark their programs against those of other universities in their state. By analyzing publicly available data from 30 of the state's other institutions, they can make faster and better operational decisions to meet student needs.

> **BYJU'S**, an Indian EdTech provider, analyzes student learning data from more than 15 million students to personalize their learning journeys and provide them with customized feedback.

> **Betha Systems**, a Brazilian technology company, works with secretaries of education in Brazil to reduce the dropout rate. Using AWS, Betha created a predictive model that is fed with thousands of data points based on 17 variables from the students' profiles. In the first year of piloting the model in the schools in Bombinhas (a city in the state of Santa Catarina), the schools were able to reduce student attrition by 20%.

Content elements:

Because these uses have financial implications—increased enrollment and operational streamlining, for example—we can think of data as a financial asset. In the commercial sector, this value is often explicit because it emerges in acquisitions and court cases, as in Microsoft's acquisition of LinkedIn, with its data on 433 million customers, for $26.2 billion, or the bankruptcy proceedings of Caesars Entertainment Operating Corp. Inc. in 2015–2017, where creditors argued that the data on the 45 million customers in its Total Rewards customer loyalty program was worth $1 billion and was its most valuable asset[6]. In higher education, data has concrete financial value as well, in addition to its mission value. Data that leads to student retention, for example, has financial value as the cost to acquire new students rises. As for the impact on mission outcomes, there is no reason to think it is less valuable than the explicit cash value in corporate acquisitions!

It is sometimes helpful to think of the business value of data as a kind of financial call option—that is, it gives us the opportunity to make changes in the supply chain or launch a new program or service but does not obligate us to do so. We can exercise the option or not, depending on how valuable the data indicates that the new initiative will be. It is here that we have had trouble finding the value of the data asset: Valuing a call option is considerably more complicated than calculating the Return on Investment (ROI) of a projected stream of cash flows. As a result, higher education enterprises may neglect the value; but as I show in my book *War and Peace and IT*[7], many of the techniques of agile IT delivery result in precisely this kind of option value.

6. Both examples from https://sloanreview.mit.edu/article/whats-your-data-worth/. A detailed analysis of the Caesars bankruptcy can be found at https://turnaround.org/sites/default/files/11.%20Paper%20-Caesars.pdf. The bankruptcy was exceedingly complex and the value of Total Rewards was included with other assets, so it is not clear what value was ultimately attached to it.

7. Mark Schwartz, *War and Peace and IT: Business Leadership, Technology, and Success in the Digital Age* (Portland, OR: IT Revolution Press, 2019).

Data, adaptability, and agility

Value is created not just by the data per se but also by the tools and processes in place to analyze it and produce those mission outcomes. In today's digital world, fraught with rapid change, uncertainty, and complexity—disruption, you might say—organizations in all industries need to use data to support organizational agility and to respond quickly and flexibly to changing circumstances. Agility is what lets organizations turn this rapid change and disruption into opportunity. Private and public sector enterprises in the digital age have learned that they can innovate and keep pace with the times through fast delivery of IT capabilities, which are then evolved and improved through multiple rounds of iteration.

The last few years have brought techniques for building agility into the IT delivery process, including agile software development, DevOps, and Lean software development. The cloud has been used to speed up the delivery of IT capabilities, for both software and hardware. Team-based organizational structures have made it possible to mobilize the resources to meet changing needs. All of these developments have helped enterprises make their processes more agile.

But agile processes are only one part of the story:

The organization's data itself must also be agile. It must be easily available for uses that are unexpected and constantly changing. It must be accessible and meaningful. Faculty and staff must have tools easily available to work with the data and the skills to do so.

It is this ability to use data flexibly—to make it available for new uses that we don't know about in advance—that is the missing link in achieving enterprise agility and distinguishes the adaptable organization from one that has merely adopted the frameworks and trappings of agile models. Business and mission agility require data agility. A data-savvy, data-informed institution is a master of both.

This focus on bringing agility to data is new. As long as data was only operational, we could lock it away in highly structured databases whose structure reflected the way it would be used for those operational transactions. Our tools were relational database systems such as Oracle or SQL Server, whose strengths are in transactional processing. We used the data to conduct the transactions themselves—registration, enrollments, etc.—and to produce compliance and operational reports.

Content elements:

To the extent that we paid attention to privacy in the past, we enforced it by strictly limiting access to the data rather than searching for ways to make it available within the bounds of privacy guardrails. Instead of "privacy by design," we practiced a sort of "privacy by obscurity." With sensitive student data in their databases, higher education institutions simply cannot rely on this sort of ad hoc, hopeful approach to protecting data.

Yes, there were attempts to free data for ad hoc analysis with so-called business intelligence (BI) systems. But the tools have now advanced far beyond what BI systems were meant to do: we now have machine learning (ML), a range of purpose-built databases to handle different types of data, algorithms for massively parallel processing, vast amounts of unstructured data like video and speech, Internet of Things (IoT) devices that deliver streams of sensor-derived data, and vast amounts of data. With these tools, we can free our data from its transactional, operational, compliance-centered context.

More importantly, we have realized that using data to inform decisions is not just a technical challenge but also an organizational one.

To be data-informed, an organization must think differently about how it makes decisions and how it interacts with stakeholders such as learners, educators, governments, and communities. It is a commitment to the value of data, a kind of organizational humility that says, "the data—perhaps—knows better than we do."

How can we make our data available to be used in new ways; that is, how can we use it flexibly, in ways we didn't foresee, to give us organizational agility? How can we apply it to bring both rigor and creativity to mission decision-making? How can the culture of higher education institutions take advantage of this new flexibility and adaptability?

And, importantly, how can we put appropriate control guardrails around the data to safeguard its privacy and at the same time allowing it to be used flexibly and quickly?

Content elements:

Agility for data – six steps

To achieve institutional agility, we'll need to be poised to respond to unexpected changes in the social and educational environments, and we'll need to create innovations that are truly novel—and so, we will need to be able to put our data to work in ways we don't necessarily anticipate when we collect it.

Our challenges:

- Our data is probably locked away in transactional, relational databases; off-the-shelf student information, learning management, enrollment, classroom management, survey data, and other IT systems; and siloed in ways that make it inaccessible to different parts of our organization.

- We may not have the right analytical tools, or they may not be available to the right people at the right times.

- Our models for security and privacy are ad hoc, as we perhaps never contemplated using the data for exploration. Most likely, we are fostering privacy simply by making the data as inaccessible as possible.

Our goals:

- Maximize the data's availability, subject to guardrails for privacy and confidentiality.

- Foster transparency across the institution by breaking down information silos.

- Offer faculty and staff the appropriate tools to explore the data in unplanned ways and in ways that take advantage of the latest advances in analytics.

- And be sure to have the expertise to interpret the data, both rigorously and creatively.

In "Analytics without Limits: FINRA's Scalable and Secure Big Data Architecture,"[8] John Brady, the chief information security officer (CISO) of the Financial Industry Regulatory Authority (FINRA), frames these objectives elegantly by saying that he wants to lower the cost of curiosity. He refers to cost in its widest sense, including the time it takes to draw inferences from the data and the risk in making it available. FINRA's business is to explore the 37 billion or more transactions that take place in the financial markets every day, looking for patterns of fraud. Since they don't always know in advance what a pattern of fraud looks like, they must rely on the expertise of their analysts to spot suspicious behavior. Their task is all about curiosity: They want their analysts to examine data with inquisitiveness as to what patterns appear and why. The task of their IT organization is to reduce the cost of that curiosity and the effort that an analyst has to exert to explore a hunch.

8. https://aws.amazon.com/blogs/publicsector/analytics-without-limits-finras-scalable-and-secure-big-data-architecture-part-1/

Brady's idea applies across organizations and roles. Can a fundraiser easily explore data to find unexpected patterns in donor contributions? Can facilities explore data to identify opportunities for lowering carbon footprint? Can finance explore data to concoct new ways to improve efficiencies? Can IT leaders test their hypotheses about how to optimize cloud spending with rigor and creativity? Can the organization explore recruitment data and student services data together with student engagement data to discover correlations between recruitment, student support, and outcomes?

Curiosity drives innovation and improvement. Agile data allows employees to freely explore ideas, hypotheses, and conjectures at the speed of thought and to promote new ideas with the data to support them.

To make data agile, an institution needs to address how and what data it gets, how it preserves that data, how and under what conditions it makes the data available, and what tools and skills it has for working with that data. Here are the six steps.

1. Get the data

To use the data nimbly, we must first have the data. And given the unknown uses to which we will put it, we need to collect more data than we know how to use. That, in a nutshell, is what "big data" is about. Fortunately, with the cloud, the cost of storing data is low and declining.

We can, therefore, instrument our business processes to produce data, lots of it, and make it available for analysis. For example, IoT applications often include sensors that blast a stream of data points into the cloud that the enterprise can analyze immediately or store away for future analysis. Enterprises can also now work with a much wider range of data types: video, text, and speech, for example. The possibilities for using all of this information in novel and interesting ways is tremendous. Higher education institutions, for example, can take various media formats of learning objects and catalog them (through tagging) so that they can be prescribed to the right learner at the right time.

> **Alef Education**, for example, a leading EdTech provider in the United Arab Emirates (UAE), collects 100 million data points each day through its learning platform to create a real-time dashboard on student progress and to trigger automated processes such as targeted support, automated assessments, and grading. This saves educators time and provides a more personalized learning experience for students.

Content elements:

2. Store it

Once we acquire the data, we must store it to make it available for analysis. Traditionally, we stored data in a structured format based on our expectations about how it would be used operationally. For example, we might have a field in a database for "course credits" and another field for "class size limit." We would collect the data to fill these fields and file them away by slotting them into the appropriate blanks in the database, knowing that we could always use these values operationally in presenting course offerings. By forcing the data into such a mold, we made it useful for transactions, but we might have lost information that could have been useful for analysis. This was the relational database model.

The past few decades have been dominated by the use of these relational databases, which are very well suited to efficient processing of old-world volumes of transactional data in ways that are known in advance ("multiply grade quantity by course credit"). But when you are working with non-transactional data or operating at tremendous internet scales of transactions or managing data that does not slot easily into pre-defined "data fields," there are now much better alternatives, purpose-designed for the cloud.

For example, Amazon Timestream is a database designed specifically to manage time-series data (like the data produced over time by an industrial sensor or by tracking market activity over time); Amazon Quantum Ledger Database is intended for the type of data used in blockchain (data whose history must be verifiable, using techniques like cryptography); and Amazon Neptune is designed for representing complex connections and relationships, like social networks. Enterprises are no longer limited to what they can force-fit into a relational model.

Better still (for agility), data that will be used for yet-undetermined analysis can be stored in a flexible repository called a data lake, where each piece of data is stored simply in the form in which it was received. The power of the data lake lies in the tools that can be used to analyze it: tools that let you combine heterogeneous information, mixing together structured and unstructured data, data from different organizational silos, and data in large quantities. Today's tools can apply ML algorithms and statistical analyses, and they can work with natural language text, video, and speech.

> **In other words, the data lake meets the enterprise need for storing data before it knows all the ways it will be used. We can pour data into the lake from different functional silos and analyze it all together.**

In the case of an acquisition (in EdTech, for example), we can quickly set up a way to pour data from a newly acquired organization into the lake and thereby gain transparency into its operations, and we can integrate its data with our own. The magic that makes this all possible is: (1) the low cost of storage, (2) the availability of tools that work with loosely structured, heterogeneous data, and (3) the availability of services that let you push data into the data lake at high bandwidth and asynchronously (just send the data toward the data lake as you receive it, and it will get there as quickly as it can, no need to wait—sort of like an email).

Content elements:

3. Make it available

The next step in bringing agility to data is to make it available—when and where it is useful. (Note that I didn't say when and where it is needed. I'm talking about agility and innovation here.) The model that is often used today is one of self-service provisioning. When an analyst at FINRA is curious, he or she can spin up a set of tools and a subset of the data to analyze without having to request and wait for someone else to provide it. The resulting freedom lets the analyst pursue a train of thought, a "flow," rather than proceeding in a stop-start way that destroys creativity—or, you could say, that increases the cost of curiosity. The cloud is an important enabler for this, as it allows new work environments to be provisioned, used, and then discarded when no longer needed. It also makes it easy to put guardrails in place to protect privacy (more on this below).

Much of the data held by higher education institutions is locked away in student information and learning management systems. The data is available for reporting within those systems—in the ways that the software vendor makes it available. To make the data truly agile, and thereby to use it as an asset for adaptability, the first step is often to move it into a data lake where it can be combined with data from other systems and made available for a broader set of flexible analytic tools, like machine learning.

> The **University of Maryville,** a four-year institution in the US, wanted to use its data to improve student outcomes but found that "in higher education data largely remains siloed into enterprise IT systems or departments. Academic or learning data and student information, such as student profiles, course completion, housing, and financial aid information, largely remains separated." They solved this by setting up a data lake on AWS and filling it with data from their siloed IT systems. In doing so, they were able to design a model that could automatically identify students who haven't activated their Maryville accounts—necessary for beginning coursework—and trigger a text reminder. The **Department of Education in Western Australia** brought together data on 320,000 students across 800 schools to analyze and serve their needs for information and communication technology devices.

Content elements:

4. Provide tools

A data-informed organization makes a range of analytic tools quickly available to its staff, often through a self-provisioning model, as described previously. A wide variety of software and services are available. If you want to perform traditionally structured queries against the data, for example, you can set up a data warehouse based on the data in the data lake, or you can provision a tool that lets you do old-school, SQL-type queries directly against the data lake.

But today, there are many more possibilities. You can, for example, visualize your data with modeling tools, and you can construct scenarios and ascertain their consequences. Today's analytics revolution is all about artificial intelligence (AI) and ML, which opens up new possibilities for what we can do with our data: predict outcomes, spot anomalies, categorize data, analyze sentiment, discover patterns, guide robots, and much more. In addition to broadening the uses of data, today's technologies can dramatically reduce the time to science.

> For example, at the **University of Adelaide** in Australia, researchers needed to analyze 48 wheat exomes and 18 whole barley genomes—some three terabytes of data. To get the needed capacity, they worked with AWS partner RONIN to create a high-performance computing (HPC) cluster in the cloud. According to Dr. Nathan Watson-Haigh, a research fellow in bioinformatics, the wheat exome analysis, which would otherwise have taken them two weeks to complete, required only six hours in the cloud. **Education Perfect** (EP), a New Zealand EdTech, uses AWS tools such as ML and rich data analytics to help instructors track student growth between pre-test and post-test periods and provide adaptive learning capabilities with real-time feedback, tailored pathways, and targeted resources. Outside of higher education, you might have seen Sky News, in their coverage of Britain's royal wedding, using AWS ML to recognize the faces of celebrities in the crowd and identify them for the TV audience, or Formula 1, Major League Baseball, and BUNDESLIGA using ML to enhance the viewer's experience. These are powerful tools that open up many possibilities for innovation in education.

To apply ML, you train a model based on earlier data sets and then apply it to new data as it is observed. In AWS, there are three general approaches to machine learning: (1) use a model such as Amazon Rekognition, which has already been trained to recognize objects in images, or Amazon Lex, which has been trained to understand intentions expressed in natural language (Rekognition, for example, was used by a number of institutions to support remote proctoring), (2) train and apply your own model based on any one of the common algorithms used for ML, using Amazon SageMaker, or (3) use your own algorithms and training approaches, if you have employees skilled in ML, by working directly with AWS services optimized for ML.

With tools such as these, institutions can unleash the creativity of their faculty and staff and find new ways to put data to use.

5. Upskill

The next important element in extracting value from your data is to make sure you have employees or staff with the right skills ... in addition to a sense of curiosity. This is why data scientists are in such high demand today. Yes, there are plenty of tools available even for people with little skill or experience in statistics. But to really make the most of data, and to do so with rigor, it is important to have people with a good understanding of how to make correct inferences from data.

For a simple example, those of us with less statistical experience tend to over-rely on averages, even when looking at an entire distribution of values can often lead to important insights. In one case I remember from my time as chief information officer (CIO) at US Citizenship and Immigration Services, we were looking to reduce the time it took us to process certain types of immigration applications. We created dashboards to track the average amount of processing time, but each change we tried seemed to have only a small impact on the metric. What we had missed was that the small number of applications that raised national security or fraud concerns took much longer to process, thereby skewing the average. We had no way to control how long those took. Although our improvements applied to the great majority of cases, because of the highly skewed average, we couldn't really see their impact. When we realized the problem and began monitoring, say, the 85th percentile completion time, we could identify the significant impact our changes had on the vast majority of cases. We had the data, the tools, and the access; we just lacked the skills to draw the correct inferences.

Data-informed decisions can also be poorly founded when the data is presented (even unintentionally) in a misleading way. In his book *The Visual Display of Quantitative Information*, Edward Tufte shows how data can be distorted or obscured by the way it is presented[9]. Again, an organization that wants to be rigorous in its use of data must make sure that it has the right skills in analysis and presentation, as well as the data.

9. Tufte, Edward R., *The Visual Display of Quantitative Information* (Cheshire, CT: Graphics Press, 2015).

Content elements:

6. Provide guardrails

Before we can make data available for novel uses—to satisfy curiosity, so to speak—we must put guardrails around it for privacy and confidentiality. Data-savvy enterprises practice "privacy by design," deliberately establishing safeguards based on planning and foresight. They gain speed and flexibility down the road by making sure that they have already considered what needs protection and have set up automated ways to protect it. In fact, the recent European Union General Data Protection Regulation (GDPR) requires privacy by design.

The cloud provides many tools for setting up automated access controls and does so at a granular level that lets you give faculty and staff access to precisely the data they should have access to. There are ways to track the provenance and validity of the data, to encrypt or obscure it, and to restrict access on a field-by-field basis or record-by-record basis.

In other words, you can specify which students' data a faculty or staff member has access to and which pieces of data associated with those students they can view.

Amazon Macie even uses ML to identify which data in your data lake is personally identifiable information (PII) and track how it is used.

Or you can choose to manage data only at an aggregated level or with information masked or anonymized. The flexibility is there; each data-informed institution must make responsible decisions about privacy given the type of data they steward.

Many other challenges arise in using the vast amounts of data that the institution has available. It is often a challenge to accurately connect data from different IT systems pertaining to a single individual, especially in countries like the US that do not have a single national identification system. Data can be inaccurate not only because of mistakes made in data entry but also because of limitations in the IT systems that collect the data. For example, there are IT systems that only allow for a surname and a given name, which imposes inaccuracy for people who have more than two names[10].

Regardless, the goal of a data-informed institution is to make data available to drive rigorous and accurate decision-making and continuous innovation. It requires collecting and storing data for flexible use later, making it and the right tools available without friction to those who will use them, ensuring privacy and confidentiality by design, cultivating the skills to make valid inferences, and solving the data hygiene problems that can lead to poorly informed decisions. This is what it means to bring agility to data.

10. For some great stories about IT systems insensitive to real-world scenarios, consult Gojko Adzic's Humans vs Computers.

Content elements:

How can we use data to bring adaptability to our institution?

Breaking down problems

Older ways of doing IT in educational institutions often led to large, monolithic, and time-consuming projects. Agile ways of working encourage breaking those down into smaller, incremental initiatives, with valuable results delivered frequently along the way. The total amount of work doesn't necessarily change (although it is often reduced, for reasons I will cover), but it's broken into manageable pieces. Doing so helps avoid the huge analysis-paralysis that accompanies large projects and not only delays them but often prevents them from getting started at all. Large, monolithic initiatives are risky and therefore require much more upfront vetting and consensus building. The smaller, more incremental approach lets institutions reduce risk, begin more quickly, get value from their efforts sooner, and constantly improve their delivery as they gain experience. It is an essential technique of today's IT delivery practices.

Working incrementally in small deliverables also vastly improves an institution's adaptability and agility. Instead of planning and committing to a project that might last five years and turn out to be outdated by the time it is finished, the institution can frequently step back and reprioritize or adjust its efforts as circumstances change. During the pandemic, organizations that had become used to working in short, incremental cycles were able to adjust to new realities as they developed. The incremental approach also frees up innovation by allowing new ideas to be tested quickly and reduces risk by allowing institutions to gather continuous feedback that they can use to adjust their direction or halt projects that don't seem to be achieving their desired effects. They can thereby always make sure they are solving the right problem at the right time.

Content elements:

Fast feedback

Feedback, in this sense, does not mean asking students or employees whether they like a new feature or IT application. More commonly, data-savvy institutions use quantitative feedback—the kind of feedback that is gathered by watching how students and employees actually act—or by monitoring educational trends or other metrics.

For example, organizations often improve the usability of their websites through A/B testing; that is, by trying two variations on a piece of the design (usually one variation is the current, status quo version, and the other is a new piece of design they are considering introducing). They show some users version A and some version B. They collect data on the users' activity and analyze it in relation to the outcomes they care about. If they want to decide whether to make a button green or red to maximize the number of times it is clicked, they can show some users a green version and some a red one and see which gets more clicks. Expedia and Netflix are examples of companies that routinely do A/B testing, drawing on large amounts of data from a data warehouse in the cloud[11].

The powerful approach of learning and adjusting through feedback goes far beyond just A/B user interface testing. Ideas for new student services or programs for non-traditional or adult learners, for example, can be tested by creating a "minimum viable product," the smallest and simplest version of the service the institution can use to gather information on whether the service will be successful or what needs to be changed to make it so. Diversity-enhancing strategies, new learning modalities, technology alternatives—all of these can be tested through trial and measurement to reduce uncertainty. And the key to doing so is gathering data and making it available for analysis.

The technique of using minimum viable products and fast feedback is described in Eric Ries's book The Lean Startup. According to Ries, at any given moment, an organization holds two hypotheses: a value hypothesis, about how their proposed product will create value for customers, and a growth hypothesis, about how the organization will be able to "grow its market"—that is, get people to use the product. The minimum viable product is the smallest product that will give the startup information to confirm or refute these hypotheses, at which point it can make changes and re-test them with the market.

This set of practices does not just apply to startups or to new product development. It has become central to the way organizations, including large enterprises in both the commercial and public sectors, achieve business agility by changing course based on their learnings. If an institution is thinking of developing a new IT system for use by its own faculty and staff, it presumably has a hypothesis about how that IT system will deliver the business outcomes that are proposed in its business case. That hypothesis should be tested, and changes should be made based on what the data shows.

As a result, agile practice requires data. To learn and adapt, an institution has to collect data on the impact of its new initiatives and use it to inform those initiatives. Agility further requires that the institution continually sense changes in its business environment, so it can respond appropriately to maximize its mission outcomes. A data-informed institution not only brings agility to its data but also uses data to support its agility.

11. https://www.youtube.com/watch?v=k8PTetgYzLA.

Content elements:

Culture and process change

Becoming data-informed, in this sense, requires a different way of making decisions; it is a deep cultural change for many organizations. In the past, we might have made decisions by crafting detailed plans, analyzing options with the available data, and choosing the option that—given only the available data—appears to deliver the best outcomes. In the digital world, we refuse to accept only the data that is available at the instant the plan is created. Instead, we design experiments to yield additional data and then incorporate that data into our decision-making. We resolve uncertainty by capturing new data.

An example is the technique for IT governance that we devised when I was the CIO of US Citizenship and Immigration Services (USCIS). Instead of writing a large requirements document and handing it over to the technologists for implementation, we simply handed over a business objective. In one case, for example, we noticed that a skilled case processor (a "status verifier") could process about 70 cases a day, and our business objective was to make that number much higher. In another business case, we found that a number of paper files got lost in transit as we moved them between processing locations, and we wanted to eliminate those losses.

For each of these objectives, we began by creating a dashboard that showed the key metric: the number of cases per day or the number of files that were missing. Instead of writing a requirements document, we created a cross-functional team of business operators and IT technologists, and we charged them with improving the metric. We gave them the tools to make changes to IT systems and business processes quickly and then monitored the dashboards with them. They tried small, incremental changes and monitored the results every day. Based on what they saw in the data, they could decide what to do next to maximize the outcome.

And management could decide whether to continue funding the initiative or direct the funds elsewhere. The result was a data-driven, reduced-risk, lightweight governance process that delivered value quickly.

This leads to another important point: accountability is enhanced by transparency. By making data widely available, we made the team's progress visible. As a result, oversight bodies could constantly revisit the investment decision, either investing more or less, redefining objectives, or stopping the investment entirely. Results were the only gauge of success, and results could be achieved quickly. But those results had to be supported by the data.

Content elements:

Spotting patterns

Another area where data can promote agility is through sensing changes or recognizing patterns in the environment. For example, ML can be used to detect and respond to anomalies. We can train an ML model with historical or routine data, so it becomes used to what is "normal" and then apply it to find activity that is not normal. This technique is used, for example, to spot fraudulent credit card transactions or equipment on a factory production line that is diverging from its normal behavior and might have to be repaired or replaced—and to do so before it actually fails. In higher education, it can be used to detect and respond to plagiarism.

> According to **Unicheck**, which uses AWS services for its anti-plagiarism software, the purpose of plagiarism prevention software is not merely to catch cheaters so they can be penalized. Instead, they want to help educators identify and find a solution for "problematic students" early on.

When we collect large amounts of data, we may find that we can identify relationships that we didn't know were there. Social media companies build large databases of relationships between people. A government might find that a potential terrorist they are investigating once lived at the same address as someone who is already known to be a terrorist—which might lead them to ask questions when they next encounter the person. A number of fraudulent immigration applications might turn out to have all been prepared by the same immigration lawyer. Here, we have moved well beyond simply using data to process transactions: we can now find important and interesting relationships between those transactions. But once again, we don't know exactly what relationships we might find; agility, flexibility, and curiosity are the keys to deriving value from data.

To cite one more example of using data to "keep an eye on events" the existence of a data point can serve as confirmation that an activity took place—for example, when audit trail logs are created automatically. By following the trail of activities, auditors may be able to validate compliance or investigate improper activity. Blockchain is often used to store data that confirms that activities took place—for example, the issuance of a credential, a transfer of money between two parties, or an approval of a contract by the parties involved. By using automated guardrails and audit data to establish compliance, enterprises can often avoid heavyweight compliance processes that reduce agility and consume precious time.

There are, of course, challenges in using data to support business agility. As we noted previously, it requires skill to draw the appropriate inferences from data. The data does not always tell us what action to take—we have to interpret it and make good decisions. Often, we face a trade-off between false positives and false negatives—for instance, if we use the data to spot anomalous transactions to identify potential fraud, we run the risk of flagging too many transactions as anomalous and annoying our customers or flagging too few and allowing fraud to sneak through. The larger the data set becomes, the more likely that meaningless patterns will emerge or that important patterns will become buried in the sheer number of potential connections. Noise accumulates along with signal.

Content elements:

In closing

A data-informed educational institution is one that puts data to work to improve mission outcomes, both by using data to drive a rigorous decision process and by making the data available for stimulating innovation and improving outcomes for students. When data is locked into an inflexible framework, siloed, or difficult to get at, it becomes a barrier to adaptability, preventing the institution from responding to opportunities or from innovating new ways to serve students. Even worse, when an organization doesn't drive its processes and investments through the use of data, it is foregoing important contact with the community it is trying to serve or passing up feedback that could help it succeed better in its initiatives. A data-informed organization, on the other hand, uses data to gain agility and uses agility to make its data more valuable. It is by doing so that we will create the higher education institution of the future.

Content elements:

About the author

Mark Schwartz is an enterprise strategist at Amazon Web Services (AWS) and the author of *The Art of Business Value*, *A Seat at the Table: IT Leadership in the Age of Agility*, *War and Peace and IT: Business Leadership, Technology*, and *Success in the Digital Age, and The (Delicate) Art of Bureaucracy: Digital Transformation with the Monkey, the Razor, and the Sumo Wrestler*. Before joining AWS, he was the chief information officer of the United States Citizenship and Immigration Service, (part of the Department of Homeland Security), chief information officer of Intrax, and chief executive officer of Auctiva. He has a MBA from the Wharton School of the University of Pennsylvania, a Bachelor of Science in computer science from Yale University, and a Master of Arts in philosophy from Yale University.

Read more from Mark Schwartz »

Learn more at
https://aws.amazon.com/education