About PoRT
Nielsen Perspectives on Retail Technology (PoRT) is a biannual publication that tracks technology trends significant to retail. It has a rotating editorship of Nielsen experts who use each edition to showcase a major technology trend in a series of in depth articles. These are accompanied by a more diverse set of snapshots on trending or interesting topics.

Executive Sponsors
Kalyan Raman, Buy Chief Technology Officer
Tom Edwards, SVP, Technology
Neil Preddy, SVP, Product Leadership

Guest Editor
Amanda Welsh, Global Connected Partner Program

Editor-in-Chief
Ian Dudley

Managing Editor
Michelle Kennedy

Associate Editor
Eve Gregg

Contributors
Kate Bae, VP, Product Leadership
Simon Douglas, Enterprise Architect, Technology Research
Sarthak Kumar, Operations & Technology Leadership Program
Ken Rabolt, Solutions Leader, Data Integration
The past 10 years have brought an explosion in the variety and scale of data available to businesses, and all indicators show this trend continuing in the years to come. As companies find competitive advantage in harnessing this data, a new reality for business is emerging. It is harder, and perhaps unrealistic, to serve the increasingly complex needs of a data-driven enterprise with a single catch-all data solution.

The variety and increasing scale of data, as well as the scope of activity it is meant to inform, demands a solution that goes well beyond a simple enterprise data warehouse (EDW). What might that more robust solution look like?

This edition of *Perspectives* features articles on three closely connected enablers that will help us answer that question.

**It is harder, and perhaps unrealistic, to serve the increasingly complex needs of a data-driven enterprise with a single catch-all data solution.**

“Your Latest Product is... an API?” explains how the ability to connect data, tasks, business units and companies through APIs is becoming a critical business competency, and one that many established firms are having to learn.

“Data, Data Everywhere, Nor Any Byte to Link” addresses the need to resolve heterogeneity in underlying data sources, where the same real-world entity may be identified and described in different ways. For connections between tasks, business units and companies to be efficient and meaningful, integrated data is vital.

“Ecosystems and the Virtual Enterprise” paints a picture for a new, optimized connection framework that allows companies to collaborate on data needs. A collaboration model distributes the burden of complexity across a network, so each participating member can better focus resources on value-added activity for their own enterprise.

At Nielsen we are investing in each of these areas. Data distribution through APIs is a key goal of Developer Studio, a critical utility in our Connected System. Data and applications in the Connected System all use normalized reference data to unify identifiers and descriptions of products, stores and markets across 106 countries. And our new Connected Partner network takes advantage of both of these resources in a burgeoning collaboration solution for the FMCG industry.

We are excited to share what we have learned on our journey. I hope you enjoy the articles and find them stimulating and helpful as you work through these questions in your own business.

Amanda Welsh
You are a refugee from a war zone without any papers to prove your identity. How do you survive in an increasingly digital world?

Bitnation—the borderless, voluntary virtual nation inspired by Bitcoin—offers a possible solution. Bitnation lets refugees prove their existence and family relations, and have this information recorded on a blockchain. The refugee can then use their blockchain ID to apply for a Bitcoin Visa card that can be loaded with funds by their relatives abroad and used throughout Europe—all without a passport or a bank account.

The blockchain technology behind Bitnation provides a way of conducting secure online transactions without the need for a governing authority. Traditional ledgers, such as those used by banks, rely on a governing authority that oversees transactions to keep them private, secure and protected against fraud and deception. Blockchains turn the principle of privacy on its head: because they are tamper-proof, blockchains can be made public without fear of fraud. And because all of the information about a transaction is publicly verifiable on the blockchain, participants can safely trade with each other without needing a trusted third party like a bank to oversee the transaction.

Clearly blockchain technology has the potential to revolutionize banking, but the retail industry is also starting to embrace it. One of the most promising retail uses is in the supply chain. Retailers need to be able to track produce as it passes through the hands of many companies, from field to shelf, to comply with food regulations, to demonstrate the validity of product claims such as “organic” and “environmentally friendly”, and so on. The London-based startup Provenance is helping companies use blockchains to document how, where and by whom their products are made, including their environmental impact. In a world where supply chain scandals, such as the unwitting use of forced labor, are depressingly common, blockchains have the potential to bolster corporate reputations.

---

1 Wikipedia provides a good explanation of the technical aspects of Blockchains.
Because blockchain does not rely on trust and authority, it promotes disintermediation—the removal of middlemen from transactions. One of the great success stories of recent years has been the rise of marketplaces created by middlemen such as Amazon, Uber and Airbnb. The blockchain lets individuals make direct, risk-free contracts for the sale of goods and services and, at the very least, could put pressure on marketplace owners to provide significant value-added services beyond risk-free contracts to justify their fees.

People were understandably skeptical when the blockchain currency Bitcoin appeared out of thin air, without the backing of a government or financial institution, in 2008. Today Bitcoin has a value of tens of billions of dollars and is accepted as a means of payment worldwide. It seems certain that the ability to have trust-free, authority-free transactions between enterprises will ultimately revolutionize far more than the world of finance.

SMART MACHINES GET SMARTER

Science fiction author William Gibson coined the term “cyberspace” in 1982, well before most people had even heard of the Internet, much less been “online”. His novels were set in a world of software yet never mentioned the programmers who wrote it, which at the time seemed to be a glaring plot hole to some tech-savvy critics. But perhaps Gibson was just as prescient about the future of software development as he was about virtual worlds.

Less than a decade ago, software developers were like ninjas writing killer code: sophisticated algorithms that did amazing things like rendering 3-D worlds and flying jet airliners. The pinnacle of this type of software was IBM’s Deep Blue, which in 1996 beat the world chess champion Gary Kasparov. Deep Blue was a programmer’s program: it took a lot of human intelligence and creativity to construct the mountain of code needed to evaluate 200 million chess positions per second and outsmart Kasparov.

Today, however, we have neural networks. A neural network is made up of many simple, highly interconnected processing elements that work in concert. It takes in data, processes it based on its internal state, and produces an output. It’s a general-purpose machine, not a program intentionally constructed to carry out a particular task; without training, it’s an empty slate. A student at Imperial College London recently taught a neural network how to play chess by analyzing five million positions from a database of computer chess games. After training, the machine was able to play at the same level as an International Master and the best conventional chess programs—all without writing a single line of code.

In order to achieve such impressive feats, neural networks need huge training datasets. They also need vast amounts of computing power so that their complex network of processing elements can formulate a response in an acceptable amount of time. Given those parameters, it’s no surprise that the leaders in machine intelligence are companies like Google, Microsoft, IBM and Facebook, which have access to both huge datasets and cloud computing power.

The potential of neural networks to solve business problems is obvious; with pioneers in the field making their machine learning libraries open-source, any company with the necessary vision and data can benefit. But not having to write software to solve business problems brings its own challenges. Unlike traditional “hard-coded” software, intelligent machines give probable answers (“these two faces are likely to be the same person”) and will make mistakes.

The models that they use to make their decisions are essentially opaque—they can’t be inspected or debugged. Making such a model part of a business-critical process will require careful design to ensure there are no unintended consequences. There are many critical problems, like handwriting and facial recognition, that are intractable to traditional algorithmic programming. But since these are relatively easy to solve using machine learning, the rise of neural networks seems unstoppable.
THE SUNSET OF THE CIO

In the next 5 years, most companies will abolish the office of the CIO and cease traditional software development.

Of course this is gross hyperbole, but there’s a grain of truth at the heart of this clickbait prediction. Changes in technology are underway and gathering unstoppable momentum, which will transform the roles of the CIO and the enterprise software development team and make them unrecognizable compared to the traditional IT department of ten years ago.

The success of cloud companies and the rise of ecosystem curators, such as Airbnb and Uber, have caught the attention of executives. To try to bring this sort of competitive edge to their own companies, they have enthusiastically adopted the agile software development practices that have served these digital companies so well. Commercial off-the-shelf software is no longer good enough.

The unpopularity of packaged software is understandable: many of these products do not support continuous integration and deployment, which are fundamental to modern software development. But cutting a lot of code to address problems that have already been solved isn’t a great idea. It doesn’t seem wise for companies to look to software development for competitive advantage when their core competencies and subject matter expertise lie elsewhere—especially in a world where technology behemoths like Apple, Google and Amazon are able to recruit the brightest and best developers. It’s hard for non-technology companies to maintain high-performing teams capable of delivering complex, cloud-architecture applications. It may also not be necessary.

The major enterprise packaged software solutions (SAP, Salesforce, Oracle HR, Office, email and so on) have already migrated to the cloud, and businesses are following them. The development of cloud platforms to meet core business needs such as customer experience, analytics and business ecosystems is undermining the rationale for having these functions in-house.

The growing adoption of the cloud and software as a service, platform as a service and data as a service, combined with the fact that these services often come pre-integrated into a digital ecosystem, greatly weakens the rationale for in-house developed or hosted software; in-house software is rapidly becoming a second-class citizen in the digital world. In an increasingly collaborative and connected world, being in the same cloud as everyone else actually enables innovation, rather than preventing it. And using the cloud provider’s infrastructure doesn’t lock you into obsolete technology: most providers are committed to offering the latest in open-source, cutting-edge software, pre-packaged for easy use.

Cloud also reduces the need for infrastructure expertise because there’s no data center to manage. The enterprise’s cloud installation undoubtedly needs careful management, but the cloud provider takes on many routine tasks. And as networks become a utility like water and electricity, working from home, on the move, in a hotel or at a hub is just as effective as being in the office; the trend towards bring-your-own device and bring-your-own app further reduces the need for a large support and networking team.

Traditional software development is disappearing as fast as polar ice. These days it’s often faster, better and cheaper for a business to source a software service they need direct from a cloud provider rather than through IT. The newest recruits are digital natives who have lived with computers their whole lives and aren’t intimidated by programming—they may not be hardcore computer scientists, but they’re more than capable of scripting a workflow and hacking together pre-existing components into a working app. The growth in marketplaces for software, analytics and algorithms will make it easier to work in this way. The ability to find, reuse and customize will become more important than the ability to cut thousands of lines of code. The convenience and agility of business users being able to assemble their own point solutions, as needed, will outweigh the benefits of a more
functional, robust, shared application developed by IT. The emergence of bots and digital assistants will also help citizen developers meet their needs without involving IT.

In the future there will be much more IT, but it will be radically decentralized. Paradoxically, this will put a higher premium on the skills of the CIO and software developers: the CIO will play a vital role in ensuring that the disparate and independent investments made by different business units can be integrated into a harmonious whole, and that the enterprise controls its costs by realizing the economies of scale and standardization. And they’ll have to balance this sophisticated governance role with the business units’ need for agility and independence. Developers will cut much less pedestrian, boilerplate code, and will instead concentrate on developing sophisticated software that’s unique to the business and provides a significant competitive edge. IT will continue to be critically important to business, but not as we currently know it.

SMART PRODUCTS FOR SMART HOMES

The Internet of Things is heralded as the next wave of digital disruption. Everyday objects are being wired with sensors and connected to the Internet to become “smart” things that can self-diagnose, detect changes to their environments and communicate with other devices via the Internet. Just as the smartphone and social media have transformed how people interact, smart things will add billions more connections to the Internet and open many new opportunities for automation.

Technology that can track consumer goods, such as RFID, has been around for a long time, but despite much effort by key industry players has failed to replace the ubiquitous barcode—the vast majority of CPG products remain “dumb” rather than “smart”. So are smart things just the latest high-tech fad, or will they have a real impact on CPG manufacturers and retailers?

The costs of sensors and Wi-Fi and bluetooth networking have decreased dramatically in recent years, but they still aren’t cheap enough that producing a host of smart CPG products makes economic sense. The drive to make smart consumer products is happening in higher-priced consumer durables: smart TVs, refrigerators, washers, dryers, dishwashers, cameras, door locks, plugs, thermostats and more. Prices will continue to fall, but—as our experience with RFID has proven—it can take a long time for this type of change to move through an industry as diverse as CPG.

Even if the food inside a smart refrigerator or the dish detergent in a smart dishwasher will likely remain dumb for some time, that doesn’t mean that the rise of smart homes won’t impact CPG manufacturers or retailers.

Once a dishwasher is connected to the Internet and a mobile phone or computer, it’s very simple to create an app that tracks how many times the dishwasher has run, how many detergent tablets it has used and when the detergent is likely to run out. The app then searches the Internet for the lowest price on the consumer’s preferred brand of dish detergent and places an order with the appropriate e-commerce site to replenish. Once the e-commerce site tells the app that the dish detergent has shipped, the app updates its inventory and starts the process again. A smart home can then let the delivery company know when people are home, or even unlock the door so that a delivery can be made.

Smart homes will create more opportunities for e-commerce to grow in the CPG space as it becomes more convenient to have smart appliances that order and replenish consumable goods. Of course this impact will vary by category; products that are a direct input to smart appliances are most likely to see changes. Smart refrigerators might even help e-commerce channels expand into perishable items, which so far have been less affected.

While most CPG products themselves will likely remain dumb for some time, the rise of smart homes—and primarily the smart appliances within them—will expand e-commerce
channels for CPG products at the expense of traditional brick-and-mortar stores.

**TONIGHT’S DINNER, BROUGHT TO YOU BY AR**

You’re sitting on your couch channel-surfing and stop on a popular cooking show. The celebrity chef is preparing a delicious-looking dish that would be perfect for dinner tonight. You only need a few ingredients—might as well grab them now.

You put on your augmented reality headset, and suddenly you’re transported to the grocery store. Your AR device is integrated with your smart TV, where a store navigation guide directs you to the first ingredient on your list. You pick up the item, check that it’s what you want and confirm that you’d like to purchase it in the dialog that appears. The shopping list floating in front of you shows a running total of the items you select as you continue shopping. Once you’ve found everything you need and paid, the grocery store begins fulfilling your order for delivery to your door.

This experience is possible through augmented reality (AR) technology, an offshoot of virtual reality (VR) that brings digital information into your environment in real time, enhancing the real world with sight, sound and more. Unlike VR—which is typically a totally immersive, isolated experience—AR lets users interact with virtual contents in the real world, and distinguish between the virtual and the real.

And now, the experience is becoming more than just hypothetical. Companies such as Microsoft, Sony, HTC, Google, Samsung and Magic Leap have already released or are currently working on their own consumer virtual reality devices. Greenlight Insights forecasts the U.S. VR market will grow to $38 billion by 2026.¹

VR and AR promise to bring revolutionary changes to both retailers and CPG manufacturers, not just consumers. Retailers would have the ability to collect very granular details about the shopping experience: they could record everything a shopper does and everywhere they look. Online retailers will also be challenged to move beyond the simple point-and-click experience on a computer monitor; the creative capacity to build out retail in the virtual world is endless.

Business analytics could also be totally transformed by inserting animation or contextual information using AR capabilities. Want to know what it would look like to add a new product line to a shelf? Want to figure out whether a promotional sign is better hanging in a window or on an end-cap? Want to see live traffic flows on a rainy day, compare them to a clear day and link it all with stores sales? AR could do all that and more, and help uncover new data points that may not have been seen before.

As AR hardware becomes more commercialized, retailers will be competing with more tech-savvy e-commerce companies in the race to bring convenient, useful and dynamic shopping experiences into the homes of consumers. Marketing campaigns will also change to become more immersive and tailored for virtual reality. And business models will also be revised, as the need for in-store employees shifts to a need for warehouse staff and “virtual experience” designers.

**THE RETAIL ROBOT REVOLUTION**

It’s a month before school starts and you’re out shopping for the kids’ back-to-school clothes, a somewhat daunting task. You walk into the store and are immediately greeted by an automated voice welcoming you and asking if they can help you find anything. You glance across the store and see a robot making its way towards you. You tell the robot you’re looking for a particular style and size of jeans, and the robot prompts you to follow it to the display where the jeans can be found. When you’re ready, you follow the robot to the checkout. The robot quickly scans the tag, asks for your form of payment, puts the jeans in a bag and thanks you for coming. Thirty minutes in the store and you never encounter a human sales associate.

As sales associates fight to raise the minimum wage, they’re also facing a new threat—losing their jobs to robot technology. The threat initially appeared in the form of self-serve checkout lines, which let customers scan, weigh, bag and pay for items without a human cashier. But replacing sales associates with self-serve checkout lanes comes at an expense. While it may be cost-effective for the retailer, there can be an adverse impact on consumer satisfaction: customers wait longer, frustrated by others ahead of them who are unfamiliar with the technology, only to face their own complications when they get their turn at the checkout kiosk (is it a single piece of fruit or a bunch? where’s the barcode? why won’t it scan?).

Thanks to advances in artificial intelligence, natural language processing and APIs, robot technology is heading to a new level of direct customer interaction and communication that’s already being tested: Amazon has started using robots to fulfill customer orders in its distribution centers; Best Buy has a customer service robot named Chloe, a robotic arm that can retrieve customers’ product selections; and in one of their San Jose area stores, Lowe’s has OSHbot to monitor inventory and help customers find items. It’s only a matter of time before more retailers enter the world of robot technology, partially for the coolness factor, but also for the potential impact it will have on their bottom line: following the initial investment, robots don’t require an ongoing wage (or wage increases), they show up as scheduled, and the only “turnover” is upkeep and maintenance.

Along with the large-scale impact on employment and customer experience, the robot revolution could also bring significant change to the world of analytics. What was once easily forgotten passing conversation between the customer and sales associate could become qualitative data recorded by a robot. Robots could help provide insight into everything from customer product requests to feedback on product placement; robots could even capture customers’ facial expressions and tone to see and hear what delights or frustrates them. Looking even further forward, our at-home Internet of Things (IOT) devices could communicate directly with these store bots, so that our favorite products are ready for purchase as soon as we enter the store. Then we could use our IOT devices to provide feedback on what we’ve brought home.

How widely will the retail robot trend spread, and how quickly? Will we ever see a store completely devoid of human employees? The answers may depend mostly on how retail technology responds to the challenge of keeping the customer satisfied.
IN FOCUS

YOUR LATEST PRODUCT IS... AN API?

As a proof of concept, Nielsen developed an Alexa Skill that let a user ask Amazon Echo questions such as, “What are the five best selling brands of tea in the UK?” Other examples of Alexa Skills include setting a timer or a thermostat, ordering a pizza and getting an Uber.

Alexa itself is not capable of doing any of these things—its skills are dependent on calls to third-party APIs that carry out the user’s request. In the Nielsen prototype, the developer added specialist capabilities that allowed Alexa to understand concepts such as brand and current period, then connected Alexa to a Nielsen Development Studio API that fulfilled the data request.¹

The Nielsen Alexa Skill may be a party trick at the moment, but the potential for a data interface that “just works” for business users, instead of them having to adapt their behavior to satisfy the interface, is too good to ignore.

The user interface is impressive, but the skill relies on the underlying API for the insight it delivers.

For a long time no one outside of IT showed much interest in APIs, but with the advent of digital that all changed. MIT research shows that the most successful digital companies make above-average investments in APIs;² these companies know that APIs are fundamental to their strategic success. Why do they think that?

WHY APIs?

Creating a single view of the enterprise from data fragmented between on-premise and cloud data centers is difficult enough. But Gartner estimates that by 2019, three quarters of an enterprise’s analytics will combine enterprise data with 10 or more data sources that belong to partners or third-party

¹ This prototype will be discussed in more detail in a future PoRT article.
² “Succeeding at Digital Requires More Infrastructure.” MIT CISR.
data providers outside the enterprise.³ These data sources include Twitter, Facebook, econometrics data, weather, market research and others.

The data lake was designed for big data but, like the data warehouse, relies on physically copying data into a centralized repository. Digital business evolves at a furious rate; new data sources become available, and old ones go extinct. Extreme big data, combined with the rapid turnover of data sources, will eventually overwhelm even the capacity of the data lake, constrained as it is by the need to copy data. APIs that connect to data remotely, eliminating the need to copy data from one place to another, are the best way to assemble a broad span of digital data in a timely and responsive manner.

Secondly, APIs have a role to play in connecting companies. It is estimated that by 2018, more than 75% of new multi-enterprise processes, such as supply chains, will be implemented as distributed, composite apps, rather than as connected monolithic applications.⁴ Integration Brokers, Integration Software as a Service (iSaaS) and Integration Platform as a Service (iPaaS) providers offer tools for wiring up the virtual enterprise, and that wiring is based on APIs. One of the market leaders, MuleSoft, has an exchange containing hundreds of connectors and APIs to help enterprises integrate their CRM, help desk, e-commerce, ERP, billing, accounting, and other systems.

Finally, APIs are the foundation of digital innovation and agility. The core data and competencies of an enterprise evolve relatively slowly. For as long as they continue to exist, banks will look after credits, debits and transactions—that's what makes them banks. But how they expose those capabilities to the world, and the way they combine them with their own and third-party services, needs to evolve rapidly. APIs are a means of exposing core capabilities like credits and debits for use in rapidly evolving technology: online and mobile banking, contactless payments, integration with blockchain-based trading systems and so on. The API both separates and connects; it lets the user experience of managing a bank account evolve rapidly and cheaply, independently of the business-critical, slowly changing core processes of the bank.

APIs allow companies to do three vital things: connect fragmented data; connect independent systems; and provide a platform for innovation—that's why companies with a successful digital strategy for the connected world are focused on APIs.

THE PITFALLS

The technical design of APIs is well understood and easy to implement. There are commercial off-the-shelf products, such as API gateways and integration platforms, which can be used to publish and manage APIs and provide security, legal compliance and privacy checks in addition to publication. These products enable APIs to be versioned and their use monitored, analyzed and regulated; they have made it easy to create a technically good API, and businesses consumed by the fear of missing out have used API platforms to rush-release APIs for internal and external use. The results have not been uniformly good.

To be successful, an API must first satisfy the developers who use it in their programs.

---


⁴ Predicts 2015: “Renovating the Core of Application Architecture While Exploiting New Principles.”
Developers are no different than their non-technical counterparts; they want an API that is easy to use. But they have a multiplicity of requirements that are not part of standard user experience (UX) design—for example the ability of the API to integrate with their favorite programming language. For this reason, some UX practitioners identify a unique problem domain, which they call “developer experience” (DX). Good developer experience, including the availability of a software development kit, documentation, sample code and so on, will do a lot to promote the adoption of an API. A vibrant developer community—that provides help, advice and best practices, as well as inspiration by showcasing innovative uses of an API—is also a tremendous asset.

But even an excellent developer experience does not guarantee success; the ultimate consumer of an API is the business, not the developer. APIs that simply expose an intelligence-free connection to a data source are unlikely to be successful. Good APIs add value by presenting data in a generally understandable form, for example based on open standards or industry norms, rather than a company’s proprietary view of the world. They make it easy to integrate data into external business processes. Passive-aggressive APIs that hand all the problems of understanding and integrating data over to the user are generally unsuccessful, but they are very easy to implement (as they require little thought) and are depressingly common. They have been shamed in the developer community as a worst-demonstrated practice.⁵

APIs are obviously not products in the traditional sense. Rather than being objects of use themselves, they are the foundation on which objects of use can be built: for example, an account management API provides the foundation on which to build a mobile banking app. But if APIs are not products in the strictest sense, it’s clear that digital businesses cannot afford to treat their creation as a purely technical challenge that’s of no concern to the wider business. To be useful, APIs must be actively managed and have a vision, strategy and roadmap, much like a traditional product.

HOW DO YOU DESIGN A COMPELLING API?

Developers used to creating legacy APIs for internal use can find it hard to re-orient themselves to create public APIs. An internal credit-check API that returned additional “helpful” but extraneous data, such as the customer’s current credit limit, account balance and debt, is relatively innocuous. But if the same API were offered publicly, it might not only violate local consumer privacy legislation, but could potentially allow the calling program to misuse the data it had been sent—and even override the result of the credit check. Returning a simple “pass” or “fail” would serve the business purpose and allow the credit check algorithm to be upgraded without any fear of negative effects in the external systems consuming the API.

Another way to create a compelling API is to capitalize the ability of APIs to deliver intelligence and business value, rather than just data. A company with an extensive database of map data could create an API that allowed them to sell maps. Third parties would be able to grab a map and build their own navigation or routing algorithms on top. The API is useful, but adds no value to the map; it would be easy to swap out if a new and better map supplier came along. Alternatively, the map company could offer an API that provides turn-by-turn directions for a journey, rather than just a map. Because the API offers directions, it’s possible to

update the map as roads get built and improve the routing algorithm without changing the API—and disrupting the software using it. The API might also encourage the creation of other services: for example, showing the location of nearby hotels and gas stations, plus the availability of rooms and the price of gas. An analysis of journeys taken using the API, and traffic information inferred from this, can be used to further improve the service. By being thoughtfully designed, the API has engendered an ecosystem.

Another pattern that has proven successful in the real world is having foundational or universal APIs and using these as the basis for creating an external API tailored to an individual client. Netflix has popularized this; they’re called “Experience APIs” because they allow an API “experience” to be tailored to the needs of each consumer.

These are a few pointers, but the use of public APIs is still in its infancy among traditional businesses. There is no proven playbook for creating digital APIs, and companies are struggling to invent solutions. Eventually best practice will crystallize and things will improve, but until then we are in for a bumpy ride. There is no doubt, however, that companies need to persist in their quest for effective external-facing APIs; as MIT has shown, that’s what the successful companies are doing. APIs are the foundation needed to connect independent systems and processes. And without the ability to connect, nothing else can happen in the digital economy.

**WHAT’S NEXT?**

Being able to connect data and processes is, of course, only the first step in understanding the environment in which the enterprise exists and realizing the value of these connections. It does nothing to resolve the heterogeneity in the underlying data sources, where the same real-world entity may be identified and described in a multiplicity of different ways and must be reconciled in order to perform any kind of meaningful integration.
In the digital economy data is fragmented, even within the enterprise. And the data that contains the most impactful insights is often outside the enterprise. Digital technology has made companies aware of the value of their own data and keen to realize revenue by selling it to third parties. Creating a marketplace where data can be bought and sold should be easy, right?

In fact it is anything but, and the difficulties are discussed in more detail in the integration and collaboration articles in this issue. Bringing disparate data sources together delivers little benefit unless they share common semantics; that is, they describe the same things in the same way. This requires independent and competing organizations to use the same semantics consistently over time—not a trivial undertaking.

Making data sources integratable is only the first step. Companies are eager to sell their data, but often the third parties most interested in purchasing it are their competitors. Providers need to be sure that the most commercially sensitive elements in their data are masked or otherwise inaccessible. For measurement companies like Nielsen, ensuring that the use restrictions stipulated by data providers are enshrined in unbreakable rules is critical to the business. Usually this is done by aggregating data in a way that makes it impossible to disaggregate to reveal sensitive information.

But this approach is not universally applicable. Many of the most analytically interesting datasets contain information about individual consumers. Thanks to recent politically charged leaks by Wikileaks and Buzzfeed, the hacking of over a billion Yahoo user accounts, and the losses of medical records and credit card data around the world, consumers are ever more sensitive about how well their information is protected and used. Realizing the value of disaggregated data while respecting confidentiality—and keeping consumer trust—is a key conundrum of the digital economy.

Sometimes it's possible to use an anonymous individual, rather than an identifiable person, in a transaction. The most obvious example is companies that target advertising at a particular consumer segment with the help of social media companies who have users' information: only anonymized or pseudonymized information is shared during the advertising purchase, so the advertiser only knows that the user corresponding to a particular anonymous “token” meets their criteria.

That's impressive, but some analyses cannot work without direct access to personal data. In these cases, data masking is one of the few viable solutions—which is why Gartner predicts that its use by large companies will quadruple to 40% by 2020. Data masking is witness protection for data: identifying data is replaced with fictional equivalents, for example by changing a person's name or credit card number. For decades companies have used static data masking, where data is overwritten in the database. But modern dynamic data-masking software intermediates data queries and fictionalizes data on the fly, making it easy to change confidentiality rules over time and have different rules for different users. It also removes the need to keep two copies of the data (the original and the masked).

Spectacular feats by academic researchers have shown that it is possible to identify individuals even in anonymized datasets. One of the best known examples is Princeton professor of computer science Arvind Narayanan's “re-identification” of individuals in an anonymized dataset. Netflix made the dataset public as part of a contest to improve its movie recommendation engine; Narayanan was so successful that Netflix cancelled a second scheduled contest.¹ This has led some companies to consider differential privacy techniques, which introduce a small amount of noise (inaccuracy) into data as a means of making it harder to re-identify individuals.²

Most enterprises simply want to monetize their own data while controlling the risk of doing so, and have easy access to a plethora of third-party data. It makes sense for these companies to outsource these problems—as they do with so many other non-core processes. All data governance problems are surmountable, and most have solutions that have long been proven in production by measurement companies. Data ecosystems, in which the curator takes care of the governance, licensing, pricing, confidentiality and other business concerns regarding a dataset, are currently one of the best models we have for doing this.

The infrastructure of the Internet is miraculous, but it only solves the technical problem of making a connection between two points. Being able to connect to a dataset may be good enough to meet the growing need of businesses to combine a multiplicity of datasets for analysis. A top-level report that puts summary company data side by side with annualized, country-level econometrics and, for example, weather, may be perfectly useful for strategic planning.

But most analyses depend on the ability to drill down into data, and to mine detailed data for new insights. The ease with which an API can be used to access data is misleading; it doesn’t mean that the data is useable. Connecting to promotional data that associates promotions with stock keeping units (SKUs) is not going to help if your sales dataset uses barcodes. Similarly, the term “data lake” gives a false impression. A lake is seamless and homogeneous: drop a bucket in anywhere, and pull up the same pure, clear water. But a data lake consists of disparate datasets—the data may be physically in one place, but logically it is still in silos.

A data analyst surrounded by incompatible data must sometimes feel like Coleridge’s Ancient Mariner, becalmed on an ocean of data but dying from a thirst to integrate it.

**INTEGRATION IS VITAL**

A category manager preparing a new assortment plan for a retailer must combine a third-party view of category sales with shopper demographics from consumer panels and the retailer’s own point-of-sale data. For such a thing to be usable, the elements common to both datasets, such as products, need to be at the same level of detail and identified...
in the same way. The data also needs to “see” the world as the retailer does—for instance, in terms of category definitions—and be augmented with metadata essential to the analysis—for example, the retailer’s policy that no more than 10% of its items have a salt content greater than 5% of the daily value per serving.

Connecting to the market, demographic and point-of-sale datasets is an essential first step, but the analysis is only possible if the datasets have been integrated so that they can be seamlessly joined and enriched to meet the business objectives of the analysis.

WHAT DOES “DATA INTEGRATION” REALLY MEAN?

Datasets are said to be “integrated” when they share common semantics. At its simplest, that means always calling the same thing by the same name. For example, an analyst needs to know that the datasets share a common definition of a store, and do not confuse it with a fascia, banner, franchise or retailer. Individual stores must be identified uniquely and universally, so that the deliveries to a store in one dataset can be connected with its sales in another.

When datasets are owned by several different companies, maintaining a common semantic model of entities, such as stores, time periods and more, is a major challenge.

However difficult this task may be, the need for integration is great and growing. As we’ve already examined in “Your Latest Product is... an API?,” the modern data environment is complex, fragmented and increasingly real-time.

According to Gartner, the number of queries accessing data sources not traditionally associated with analytics or data warehouses increased by 40% between 2014 and 2015.¹ Legacy sources of information about product purchases such as electronic point-of-sale systems are still crucial, but need to be supplemented by information from e-commerce sites, Dash buttons and the like to get a complete picture. Similarly, new payment methods such as Google Pay, Apple Pay and Amazon Payments need to be considered alongside legacy credit card data.

Digital retail has led to an explosion in the amount of contextual information about purchases: from user profiles, beacons, sensors and the Internet of Things (IoT) to advertising views, review and recommendation sites, social networks, weather and econometrics. APIs, data federation tools and plain old data copying make it easy to bring this data together, but unless the disparate sources have been integrated based on a common set of semantics, it will be difficult to mine the combined data for insights.

DATA GOVERNANCE 2.0

Integrating disparate datasets according to a common set of semantics is exactly what the Enterprise Data Warehouse (EDW) was designed to do. Master Data Management tools, governance processes, data experts and guidelines such as the golden record and “single version of the truth” generally

ensured that the data in the EDW was complete, accurate, integrated and subject to appropriate access controls. And because the EDW was a single, logically centralized repository, data problems were easy to diagnose and rectify. The downside of the EDW approach was that managing data this way was time-consuming, required a considerable number of subject matter experts, and generally only encompassed data within the enterprise.

It is accepted that EDW processes cannot scale to meet the needs of digital business;² the volume, variety and velocity of digital data is too great. Forrester has identified a modern alternative to the EDW approach, which it calls data governance 2.0: “An agile approach to data governance focused on just enough controls for managing risk, which enables broader and more insightful use of data...”³

The needs of governance 2.0 are similar to those of governance 1.0. However, datasets need to be integrated by a combination of key and text matching. They need to be normalized, for example by reconciling data at different periodicities (daily, weekly, monthly, etc.). And they need to match the enterprise’s view of the world—its taxonomies, schemas, and mental models.

The amount of data to be integrated and the speed required means that a lot of this work has to be done by machines, albeit supported by humans. But even with automation, the time and effort required is still so great (and costly) that it is critical to match the investment in integration to the business benefit. Rather than trying to create perfectly integrated data capable of serving any and every purpose (the EDW approach), integration just needs to be good enough and quick enough—but no more. Integrating data at the most detailed level possible and enriching it with complex taxonomies maximizes its usefulness for analytics and its conformity to the enterprise’s view of the world.

Rather than trying to create perfectly integrated data capable of serving any and every purpose (the EDW approach), integration just needs to be good enough and quick enough—but no more.

RISK, AGILITY AND THE IMPORTANCE OF TRANSPARENCY

A “good enough” approach to data integration driven by a “just do it” mentality is not without risk. Datasets are often incomplete and inaccurate. And even when they are sound, two datasets may represent different portions of the real world: for example, one may contain point-of-sale data for all shoppers, and the other advertising exposure data for millennials only.

If the data is required for business-critical decisions of high impact, it may be cost-beneficial to address these issues of accuracy, completeness and compatibility just as we would in an EDW; however this would greatly compromise agility. Data only needs to be good enough for the analytic purpose for which it will be used. Sometimes all that is needed is a directional read on a market, and it may be possible to get this even from incomplete, inaccurate and unrepresentative data.

To avoid problems, two things are critical in this sort of analytic environment: transparency and risk management. In the governance 2.0 world, data transparency holds the prime position that data quality and completeness does in the EDW world. Transparency means that all the characteristics of a dataset—its accuracy, completeness, representativeness, adherence to standards, the conditions and constraints under which it is licensed, and so on—are freely visible to anyone entitled to access the data.

Data transparency allows data scientists, or even bots, to make an informed decision about the utility of a dataset for a particular analysis, along with any inaccuracy or bias that might result from its use. With this information, they are able to put in place an appropriate strategy to mitigate the risks involved in using the data. Unlike the EDW approach, which tries to predict and control every possible usage of the data and eliminate risk, Data Governance 2.0 puts the power—and also the responsibility—in the hands of the data user. This enables agility, but at some risk: organizations might be willing to give great latitude to data scientists with the expertise to make well informed choices, but restrict use by less well informed citizen data scientists.

**THE FUTURE**

Traditional data warehousing initiatives by large enterprises involved them “going it alone”: assembling and harmonizing all the data they needed internally with limited help from third parties, except as subcontractors. Like so many legacy processes, this is not fit for purpose in a digital economy; what worked for the EDW in a world with far fewer and less diverse datasets does not scale for big data.

Digital business, access to ecosystems, and the benefits of network effects depend on companies in an ecosystem being open and sharing, and adopting a common purpose and standards. Most enterprises have a lifetime’s experience of proprietary models that have up until now been a clear source of competitive advantage. It will be hard, not just technically but also in terms of company culture, for companies to be open enough to realize new digital opportunities without compromising their competitive advantage.

Clearly, companies are going to have to change their behavior considerably if integrating datasets is going to become as easy as snapping together Lego bricks. Intermediaries such as data brokers and integration brokers offer one possible solution, but this is essentially outsourcing the problem. Fundamentally solving this problem will require the creation of communities of interest responsible for data governance, or the evolution of data and analytic ecosystems. A business *lingua franca* of common semantics that can be supplemented by proprietary add-ons seems like the best solution, as discussed previously in *Perspectives on Retail Technology*. 
A company may look like a single entity, but scratch the surface and you find that it’s dependent on subcontractors, service providers, partners, vendors, consultants and many more external entities. Despite outsourcing some functions, traditional companies are generally far less ruthless in concentrating on their core competencies than those that were born digital: Netflix lives and dies by its ability to deliver digital content to its customers, but owns no technology infrastructure and relies on Amazon’s cloud.

Born-digital startups can rent web-scale, state-of-the-art infrastructure from cloud providers at the swipe of a credit card and instantly offer a quality of service and geographic reach it used to take decades to build. Easy access to cloud-scale commerce is one reason why companies joining the FMCG supply chain are now on average 30% smaller than 5 years ago. The desire of consumers, especially millennials, for custom, artisanal and authentic products from smaller, specialized, innovative businesses is another contributing factor.

The nature of the enterprise is also being affected by the behavior of the always-connected consumer. Today’s shopper has more complete information about products and more ways to interact with companies, brands and services than ever before. This seamless connectivity is influencing consumers to value experience above the product: a pasta brand that provides recipes and “how to” videos may be more attractive to consumers than one that does not, regardless of the relative merits of the products.

All of these trends mean that enterprises are becoming more complex and interconnected and have less well-defined boundaries, at least as far as consumers are
Its success will be determined by its ability to coordinate a swarm of ever-changing collaborators and insulate itself from their failures, rather than building decades-long relationships with a handful of partners.

concerned—consumers don’t differentiate between a problem with Netflix itself and one with its underlying Amazon infrastructure. Gartner provides further evidence for this transformation, estimating that by 2018 more than 75% of new multi-enterprise processes will be implemented as distributed composite applications, rather than as connected monolithic applications.¹

It’s no surprise that analysts have abandoned the idea of simple, linear, value chains and talk about value webs instead. The enterprise is going virtual and digital; its success will be determined by its ability to coordinate a swarm of ever-changing collaborators and insulate itself from their failures, rather than building decades-long relationships with a handful of partners.

INTEGRATION SERVICES

One of the key competencies of the virtual enterprise is integration, and many software services have arisen to help satisfy this need. Integration Brokers (IBs) provide services that facilitate integration between systems at the application layer (as opposed to the data layer or in the user interface). They are used to integrate previously independent applications or services—within or outside the enterprise. IBs add value by taking care of the technical complexity of the connection between systems (for example, by transparently handling the different communication protocols and data formats used by the systems being integrated).

Integration Software as a Service (iSaaS) and Integration Platform as a Service (iPaaS) provide additional functionality and productivity by offering a suite of cloud services that allow companies to design, develop, run and govern integration flows connecting any combination of on-premise and cloud-based systems. iPaaS platforms de-skill the process of integration to the extent that it can be done by a non-specialist developer, or even an end user. They do this by offering preconfigured integration endpoints and ready-to-use integration flows for existing software-as-a-service applications such as Salesforce (CRM), ServiceNow and Zendesk (help desk), Shopify (e-commerce), Intuit and Square, and other ERP, billing and accounting services.

INDUSTRY PLATFORMS AND COMMUNITIES OF INTEREST

Integration services provide excellent switchboards for connecting systems with minimal effort—a necessary foundation for creating a virtual enterprise. But a great deal more, such as data and process integration, is required to create a fully functional and agile enterprise. There’s need to go beyond simply making connections to having common standards and processes. This is one reason why IDC predicts that by 2018 more than 50% of large enterprises—and more than 80% of enterprises with advanced digital transformation strategies—will create or partner with industry platforms.²

The good thing about industry platforms and communities of interest is that their members have common problems and a mutual interest in the strength and well-being of the community. Being part of an integrated industry platform takes away the need for many-to-many point integrations by routing everything through a single industry hub.

One of the best examples of a community of interest is GS1, an international non-profit organization focused on product identification and the taxonomies needed to support a supply chain. The Global Data Synchronisation Network (GDSN) arm of GS1 allows trading partners, such as suppliers and retailers, to share information about product data. GS1 maintains a central directory of “data pools” of product information, which it ensures comply with the GS1 standards. Organizations wishing to participate in the GS1 community of interest join a data pool. Suppliers and retailers can update information about their products in the pool and this information is automatically shared with their registered partners, rather than the owner needing to publish the update to each of its partners. This is an obvious (if simple) model for other communities of interest that could reduce the friction of the many connections needed to be a successful virtual enterprise.

ECOSYSTEMS

The step up from industry platforms is to ecosystems. An ecosystem is a living network of organizations—including suppliers, distributors, customers, competitors, government agencies and so on—involved in the delivery of a specific product or service through both competition and cooperation.³ What makes ecosystems unique is the diversity of their membership, their ability to foster both collaboration and competition, and their ability to innovate much faster than traditional associations.

If enterprises are to integrate data, processes and systems, they will require some form of standards and governance. Heavy-handed, top-down governance of the sort traditionally seen in the enterprise data warehouse world is unlikely to work. Although successful ecosystems have a governance framework, their success seems to come in part from the fact that they are largely self-organizing and self-regulating.

Open-source software provides a proven model for one possible approach. It has been adopted by the Global Food Safety Initiative (GFSI), a non-profit organization whose members include many of the world’s largest food producers, distributors and retailers. Shared certification, standards and monitoring, combined with the free dissemination of best practice, helps improve the safety of the food industry overall and boosts consumer confidence. Every participant benefits from the collective investment in common resources, and better food safety confers the equivalent of herd immunity against reputational damage. As the GFSI says, “Food safety is not a competitive advantage;” it is a common good that creates a foundation on which manufacturers and retailers can continue to innovate and compete.

Blockchain demonstrates a different approach to peer-to-peer and authority-free interactions between enterprises. Transactions on a blockchain are public, transparent, trust-free and cannot be falsified. It seems an ideal technology for the value-web world of increasingly numerous and complex interactions between enterprises. Perhaps, for example, the blockchain could be used as a governance mechanism for resources such as the metadata needed for data integration.

It is likely that different approaches will be best suited to different use cases.

THE VALUE OF A CURATED ECOSYSTEM

Although ecosystems suffer if there is heavy-handed governance or other unnecessary interference in peer-to-peer interactions, ecosystem owners can add a great deal of value through active curation. Promoting the ecosystem improves the community for all members, as does gap filling (for example in membership, data and functionality) and helping to set standards for data formats and content, which aid integration.

Ecosystem curators can take care of mundane but difficult and time-consuming tasks, like agreeing to the licensing of data and ensuring that conditions of use are followed, including regulatory compliance. This removes a burden from

³ “Business Ecosystem.” Investopedia.
the users of the ecosystem, who only need to have a single legal relationship with the ecosystem curator, rather than individual relationships with every other member.

Curators can provide a framework for monetization of data and analytic assets. This could include flexible pricing modeled on the Uber approach of burst pricing: owners might decide to make their data and functionality free for academic use (as a spark for innovation), but value-priced when used by their peers. Curators can also act as an intermediary between sellers and buyers: the curator could be given the responsibility to “protect and share” the data, preserving the data owner’s confidential information by aggregating or anonymizing it before it is shared, or by running analyses on behalf of the data consumer and providing summarized, non-disclosing, results.

One of the biggest advantages of curation is that it can help participants realize what they don’t know. A complete, detailed and searchable catalog of the assets in the ecosystem is table stakes. In a successful ecosystem, the catalog may be as full of things of uncertain usefulness as an app store. Microsoft’s Azure Data Catalog uses the TripAdvisor approach to solve this problem, allowing consumers to comment on and rate the assets they use. If there is a sufficiently large, motivated, objective and well-informed user base this could be very valuable, but there is currently limited evidence that this sort of approach works as well in the B2B world as it does in B2C. In a B2B ecosystem there may be a great deal of value in the curator actively filtering, prioritizing and editorializing the content shown to a user based on their needs and preferences. There is definitely a role for an assistant, probably using machine learning, to bring useful information—unobtrusively—to their notice.

The curator can also protect users from the high degree of turnover of startup companies, and the rapid evolution of the data landscape and big data technology. A safe, curated space that provides access to all of the latest and greatest data and analytic “toys” adds a lot of value. Clients don’t need to worry about chasing trends in new technology, they can simply use it as it appears in the ecosystem.

All companies are seeing IT spend drift away from the control of IT and business units becoming autonomous. That can’t be controlled, but committing to an ecosystem (or even a data governance framework) will mean independent investments are interoperable and synergistic.

One of the biggest advantages of curation is that it can help participants realize what they don’t know.

...committing to an ecosystem (or even a data governance framework) will mean independent investments are interoperable and synergistic.

LOOKING FORWARD

We are still at the beginning of this journey. Enterprises could continue to assert their independence and isolation and preserve their historic, proprietary competitive advantage. But the story told by open source and the rise of digital businesses and ecosystems suggests that companies are more likely to succeed when they take common-cause foundational capabilities and build differentiation on top, rather than trying to create proprietary solutions from the ground up. This is difficult to do and, outside of the born-digital companies, most enterprises are only taking their first steps.
Cloud integration platforms, industry platforms and communities of interest are already playing a part in building the virtual enterprise. But by and large, they simply help to connect existing capabilities. In theory, industry platforms have the potential to assemble a critical mass of committed parties and experts, and as a result they are likely to succeed. Their weakness, if they have one, is that they are predicated on a particular view of the world: the existence of the industry and its legacy processes, products and services. Digital has revolutionized traditional industries like hotels, taxis and retail, and the transformation shows no signs of slowing down. It doesn’t seem like a good idea for an enterprise to align itself too closely with a particular model of an industry—which could be rendered obsolete by innovation.

Ecosystems have a proven track record of remarkable agility and innovation that is hugely appealing; however we’ve yet to identify a simple formula for creating a successful ecosystem. Ecosystems cannot exist without a degree of sharing and openness, but to attract participants they need to exist at the equilibrium point between the benefits that arise from openness and common standards and the competitive advantages that come from proprietary mental models. In 2013 Microsoft set up their Azure Marketplace for buying and selling Software as a Service applications and premium datasets. They recently announced the retirement of the Marketplace DataMarket and Data Services with effect from Q1 2017. Curating an ecosystem is not a trivial task.

Nielsen believes that ecosystems built around competencies, rather than general purpose or industry ecosystems, are the most likely to gather the critical mass and subject matter experts necessary to sustain their existence; enterprises should join ecosystems corresponding to their core competencies, such as supply chains. In addition, they’ll need to be members of ecosystems that provide the infrastructure they need to conduct their business effectively, such as analytics and market research. Defining the enterprise as a set of competencies and interests—instead of aligning it to a particular industry model—will allow it to evolve rapidly by joining and leaving the appropriate communities, or even creating new communities, as changes in its business environment dictate.

As the business landscape increasingly organizes itself into dynamic, interactive ecosystems, value webs will evolve. Larger firms will invest in their own ecosystems, recognizing that feeding and nurturing them will help generate demand, innovation, and support in a variety of ways that cannot always be predicted. New leadership capabilities will be increasingly valued, as relationships based on reciprocity, mutual trust and shared interests become even more important.

ABOUT NIELSEN

Nielsen Holdings plc (NYSE: NLSN) is a global performance management company that provides a comprehensive understanding of what consumers watch and buy. Nielsen's Watch segment provides media and advertising clients with Total Audience measurement services for all devices on which content—video, audio and text—is consumed. The Buy segment offers consumer packaged goods manufacturers and retailers the industry's only global view of retail performance measurement. By integrating information from its Watch and Buy segments and other data sources, Nielsen also provides its clients with analytics that help improve performance. Nielsen, an S&P 500 company, has operations in over 100 countries, covering more than 90% of the world's population.

For more information, visit www.nielsen.com.

Copyright © 2017 The Nielsen Company. All rights reserved. Nielsen and the Nielsen logo are trademarks or registered trademarks of CZT/ACN Trademarks, L.L.C. Other product and service names are trademarks or registered trademarks of their respective companies.