GETTING READY FOR PHARMA 4.0
Data integrity in cloud and big data applications

Toni Manzano, PhD, and Gilad Langer, PhD

The amount of data collected in a typical pharmaceutical manufacturing operation is staggering, yet research shows that much of this information is rarely used for anything more than compliance. New technologies such as big data, artificial intelligence, machine learning, and deep learning permit unprecedented analysis of real-time data and can even predict trends in processes and operations. Manufacturers can use these technologies and the information they provide to understand and improve their processes.

Machine learning (ML) can be described as a way of achieving AI through "brute force"—superfast, relentless calculation that gauges every possible option in search of a solution. Deep learning (DL) uses algorithms based on the structure and function of human learning to cascade and transform data through layers of processing. Big data technologies can store and retrieve huge volumes of data at high speeds. And artificial intelligence (AI), which can learn human activities such as planning, language comprehension, and problem-solving, offers advanced analytics to produce meaningful conclusions and predictions from these data sets. Together, these technologies deliver the information and intelligence needed for continuous improvement—one of the promises of Pharma 4.0.[1] But with them comes a question: How can we maintain data integrity, especially to support GMP operations?

Big data environments and their algorithms must be designed to follow data integrity guidelines. This requires a clear and well-coordinated effort to apply best practices to system design, including system architecture, data capture and storage, and data consumption. Of greater interest, however, is that these technologies permit a built-in, automated audit trail. This allows AI to analyze captured data and trigger alerts about noncompliance or data integrity issues.

Current pharmaceutical manufacturing has varied systems to manage GxP tasks, as well as the required data capture capability for analytics and real-time manufacturing intelligence. Yet research shows that 70% of all manufacturing data collected is not used, and that pharmaceutical manufacturing data capture operations in general have significant waste. Faced with increasing pressure to optimize, many manufacturers see the advantages of using already captured data to gain insights about processes and operations. The ability to use this data, however, given the multitude of monolithic systems, each with their own proprietary format, is not trivial. The
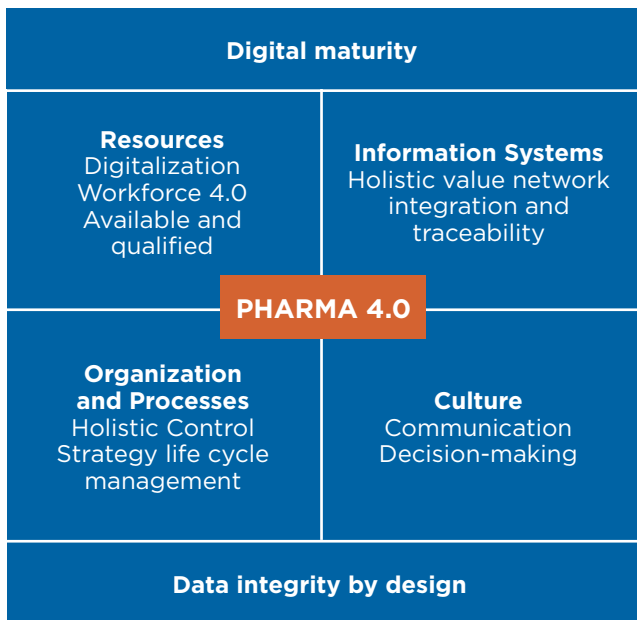
Industry 4.0 paradigm promises to solve many of these shortcomings with technologies such as big data and the Industrial Internet of Things (IIoT).

INDUSTRY 4.0

Industry 4.0, dubbed the Fourth Industrial Revolution, was first presented in 2011 at the Hannover Fair. This new paradigm, which began as the German government's technology strategy, applies science- and risk-based approaches to manufacturing and process intelligence. An ISPE Special Interest Group has redefined it as "Pharma 4.0" for pharmaceutical manufacturing. Figure 1 illustrates its four main principles,[1] which demonstrate the importance of data and manufacturing intelligence.

Transforming current pharmaceutical manufacturing to Pharma 4.0 requires a new approach to manufacturing and process data capture. Big data provides a practical solution that does not require complicated integrated models and permits the use of cloud-based advanced analytical techniques for artificial intelligence. With enough information and quality content, these technologies can transform data into knowledge that supports critical activities, including process optimization, continuous improvement, operational excellence, and

....

FIGURE 1: PHARMA 4.0

Digital maturity	
Resources Digitalization Workforce 4.0 Available and qualified	**Information Systems** Holistic value network integration and traceability
Organization and Processes Holistic Control Strategy life cycle management	**Culture** Communication Decision-making
Data integrity by design	

PHARMA 4.0

TABLE A: COMPONENTS INVOLVED IN CLASSIC IT AND XaaS SCENARIOS

	Iaas	Paas	Saas	Classic IT
Network	■			■
Storage	■			■
Servers	■			■
Virtualization		■		■
Operating system		■		■
Middleware		■		■
Data			■	■
Application			■	■

even real-time release. Pharma 4.0 also allows the potential for GxP decisions based on the results of AI algorithms. This makes compliance with data integrity guidelines critically important, specifically those from the US Food and Drug Administration (FDA);[2] the European Medicines Agency;[3] as well as other relevant regulations, such as the US Code of Federal Regulations Title 21, parts 11,[4] 210,[5] and 211;[6] and the EudraLex Annex 11.[7]

Other industries' progress toward Industry 4.0 has proven the value of a fully connected environment. It has shown how big data analytics can optimize operations and improve quality. The pharmaceutical and biotech industries should be able to reap these same benefits—as long as they can establish data integrity and manage it in a compliant manner.

WHY BIG DATA?

In Industry 4.0, manufacturing environments are fully connected, with every operation and piece of equipment transmitting data in real time. This includes the full plant, from devices to operational systems, across the entire manufacturing operation (work center, process cell, production unit, production lines, etc.). As a result, the amount of data collected is enormous and varied, from time series process data to complex data sets such as batch records.

Even a medium-size facility can collect between 500 terabytes to 10 petabytes of data per year. (As a means of comparison, 1 petabyte is four times the content of the US Library of Congress.) Although these numbers seem enormous, they are normal in a modern facility and will only increase as new equipment, systems, and devices are introduced. For this reason, a big data solution is necessary for storage and indexing, so that information remains accessible for historical and real time analysis.

Managing this amount and variety of data is not a trivial task, but traditional manufacturing intelligence systems are not able to do it in an effective way. The cost and effort required to maintain this volume of in-

formation along the entire data GMP workflow (acquisition, access, backup, retirement) can quickly become unmanageable. These increasingly large amounts of data require a substantial investment in data centers, backup infrastructures, and IT services. As data volume increases, periodic upgrades will incur additional costs.

But there's an alternative to these on-premise architectures: cloud-based big data services.

WHY THE CLOUD?

Cloud-based options depend on the services required:

- Infrastructure as a service (IaaS) outsources the physical hardware and the logic necessary to maintain it. Computing resources such as servers, hard disks, and the tools to manage them (regional location, data partitioning, scaling, security access, storage life cycle, and backup) are offered as a service.
- Platform as a service (PaaS) enables software development and deployment without the need to buy it. Companies such as Google, Microsoft, and Amazon are offering these types of ready-to-use components.
- Software as a service (SaaS) runs software applications on a cloud-based platform; no software is installed locally. This is very different from traditional on-site applications.

These "XaaS" alternatives provide new methods that can meet industry requirements for large data storage and computing resources at a reasonable price. They open a new way to use big data with global access, built-in security, and an inherent audit trail. They provide the necessary computation power and data storage to process huge amounts of data with AI algorithms, because they were developed for use with massive amounts of information, a critical requirement for Pharma 4.0.
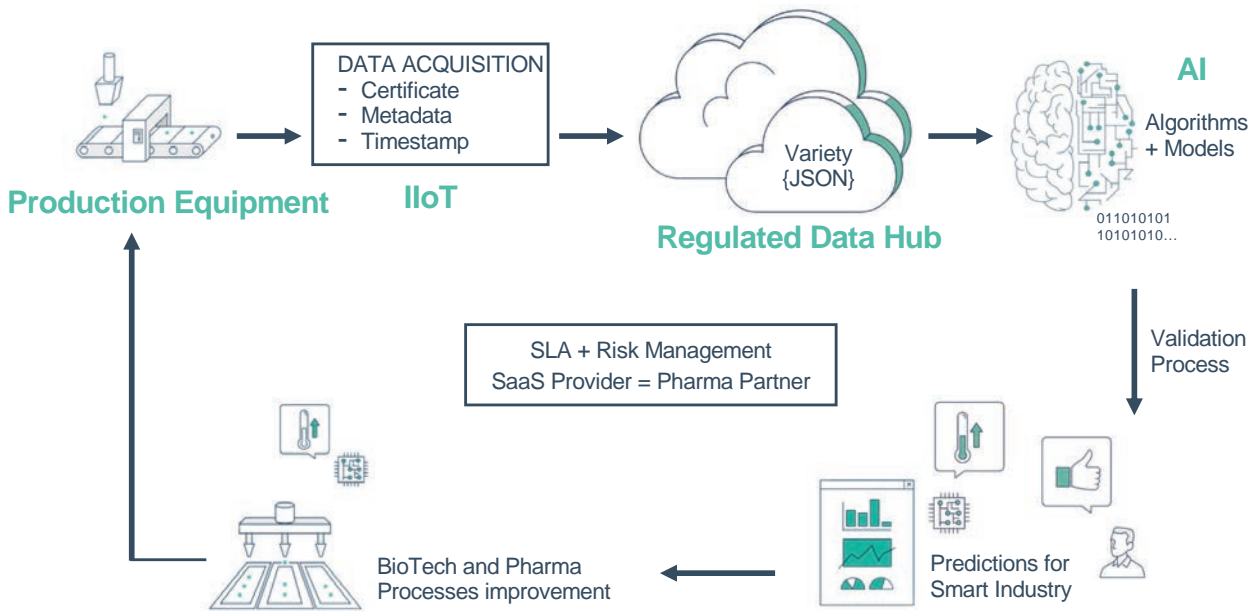
WHAT ABOUT BIG DATA AND AI?

Simply capturing data does not in itself give us knowledge. Data must be processed into information, which in turn is transformed into knowledge. This requires context and analysis:

- Big data keeps getting bigger. In the past, a terabyte was considered big data; now a petabyte is the norm.
- With growing volumes of data, classical statistical methods such as the student's T distribution, chi-square, and analysis of variance are not practical.
- The data is unstructured, with multiple formats (numerical, categorical, visual, document) that make traditional analysis methods impossible.
- The processing power required to analyze terabytes and petabytes of data is practically impossible to achieve in traditional server-based systems.

AI models, however, can analyze these huge volumes of nonstructured data using algorithms such as K-Means, Random Forest, or K-Nearest Neighbors. While these methods are not new (they were developed during the 1960s), they become incredibly useful when combined with modern cloud computing power and a vast amount of data. These three elements (big data, cloud computing, and AI) present a new way to apply science to understand the complex nature of modern manufacturing. They provide the ability to monitor production processes where equipment, devices, processes, systems, and

FIGURE 2: DATA AND DECISION FLOW IN BIG DATA ENVIRONMENT

DATA ACQUISITION
- Certificate
- Metadata
- Timestamp

Production Equipment

IIoT

Variety {JSON}

Regulated Data Hub

AI

Algorithms + Models

011010101 10101010...

Validation Process

SLA + Risk Management
SaaS Provider = Pharma Partner

BioTech and Pharma Processes improvement

Predictions for Smart Industry

THE AMOUNT OF DATA COLLECTED IN A TYPICAL PHARMACEUTICAL MANUFACTURING OPERATION IS STAGGERING

operators are continuous data sources. They also set the stage for continuous process validation by making it more practical, achievable, and cost-efficient.

Current best practice pharmaceutical manufacturing analytics use statistical methods on specific control process parameters (CPPs) and control quality attributes (CQAs) to monitor product quality. In reality, however, these parameters are dependent on external and internal factors as well as their own behavior. To truly understand product quality, we need to adopt a holistic view that considers not only CPPs and CQAs, but all aspects of the manufacturing environment along with the inherent interdependencies of the process variables.

Big data with AI can support this holistic view. Their ability to find pat-

terns and dependencies beyond those that traditional statistical methods can supply provide an approach so powerful that they have been adopted in many manufacturing industries. The pharma industry already considers AI a valid tool to manage data in its research, development, and manufacturing processes. Both artificial neural networks and support vector machines have recently been proposed by the European Pharmacopoeia as valid chemometric techniques for advanced analytical methods.[3]

Until now, great effort has been expended to provide a common context—an integrated information model—for manufacturing intelligence solutions; this is typically based on International Society of Automation hierarchies ISA-95 and ISA-88. As noted previously, however, data stored in a big data environment is unstructured. Whatever context existed in the automation and manufacturing systems, therefore, is lost.

But unstructured data also means that there is no need to develop an integrated model. In fact, AI algorithms not only perform better with unstructured data, they can process data from multiple sources in different formats. Because context is still important to interpret the data and create AI models, future manufacturing intelligence solutions must be able to import contextual hierarchies from automation and manufacturing systems and build it as an overlay on the big data. These overlays do not restructure the data (in storage), they simply point to and set boundaries on the data. Overlays can provide multiple perspectives on data sets such as master recipe model, equipment model, product model, etc. It is remarkable, for example, how running AI algorithms to uncover causalities will detect patterns that reinforce traditional hierarchies and sometimes even point to relationships that would otherwise have been missed.

FIGURE 3: WARNING LETTERS THAT REFERENCE DATA INTEGRITY ISSUES FOR AMERICAN AND INTERNATIONAL SITES, 2013–2017
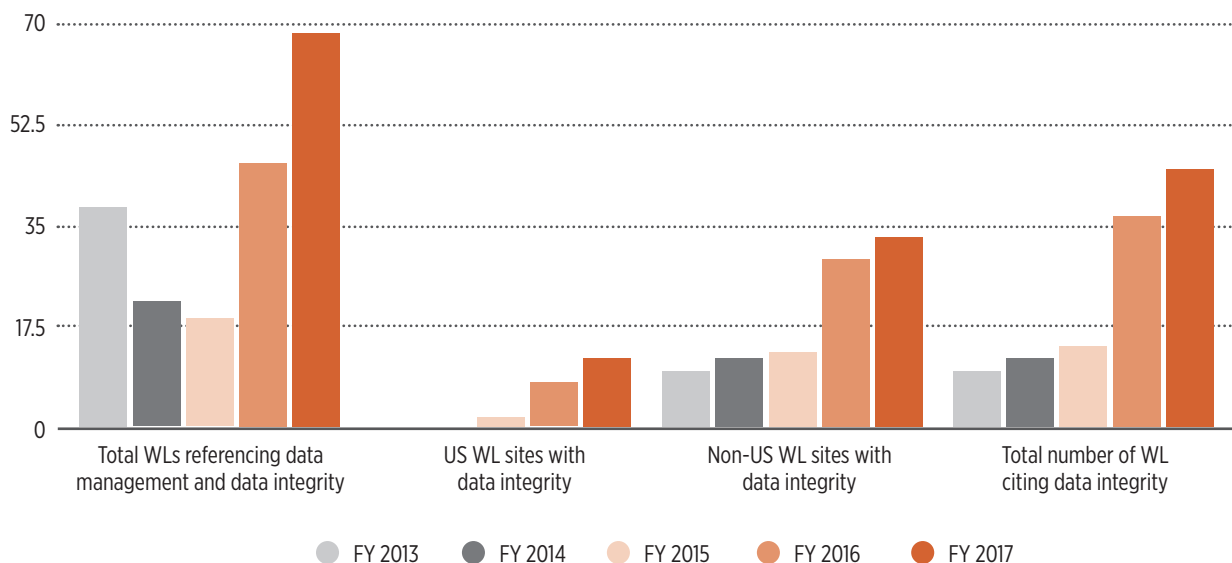
QUALITY AND SECURITY

SaaS and cloud-based solutions introduce a new paradigm for manufacturing information systems. Because there is no need to deploy and maintain on-premises IT infrastructures such as storage, servers, virtualization, operating systems, middleware, data, and applications, there is virtually no cost associated with deploying and maintaining the IT infrastructure. Table A compares the responsibilities of each component in the different scenarios.

Traditional IT applications, system architecture, data, and software are deployed on-premises, i.e., in a company's IT physical infrastructure. In a XaaS environment there are no physical servers or storage devices; you cannot obtain a server's serial number, and there is no need for a backup process. More importantly, the quality of the data that is captured, stored, and consumed cannot be measured or managed using the same classic IT systems approaches. Cloud computing technology can provide any required storage and computing capacity instantaneously. This concept, called "elastic computing," offers unlimited storage and automatic scalability.

XaaS also introduces competitive elements such as pay per use, business focus, robust systems, well-protected infrastructures, automatic backups, and total workflow encryption. Current estimates predict that by 2021 more than half of the global companies currently leading in the adoption of cloud solutions will have moved all of their systems to a cloud-based infrastructure.[8] Cloud-based solutions have demonstrated robustness and security that are far superior to traditional hardware and software designs. This is confirmed by the rapid adoption of these systems in a variety of industries such as finance, automotive, and health care—sectors that require a high degree of confidence in the technological tools.[9]

TABLE B: ALCOA+ TECHNOLOGICAL CONCEPTS THAT WORK AS DATA INTEGRITY FACILITATORS

Attributable	PGP, certificates, MFA, geolocation, time stamp, blockchain
Legible	JSON, OCV-OCV, AI (sound, image, and text recognition)
Contemporaneous	NTP synchronization, MQTT protocol, 5G wi-fi, high data speed
Original	Data variety, encryption, self-replication across regions
Accurate	Metadata, regional settings detection, primary data acquisition
+ Completeness	No limits about data volume, unstructured information, VPC
+ Consistency	Automatic audits, full traceability, electronic inviolability
+ Enduring	Replication, persistence guaranteed by big data design
+ Availability	Internet bandwidth, geographic distributed cloud, multiplatform

Compared to other industries, however, the adoption of cloud-based solutions in the pharma and biopharma industries is lagging, mainly because of concerns related to security and quality. Qualification of these solutions (IaaS, PaaS, or SaaS, or a combination) requires a risk-based approach and reliance on the service provider's quality systems and processes. This often challenges the traditional qualification mindset.

There is no physicality to these solutions; no system hardware, no server instances, no serial numbers, and no operator systems to verify. Operational aspects, such as the management and compliance of data and information storage or backup processes and their storage, should be delivered by the PaaS provider. Data storage and computing systems are provided on demand in a "serverless" environment. Disaster-recovery plans become simpler and faster with architectures that simultaneously replicate data and information in real time to different geographic locations ("geolocation replication").

SERVICE LEVEL AGREEMENT

In this type of environment, where hardware and operational functions are delegated to the XaaS supplier, the service level agreement (SLA) is the end user's—in this case the pharmaceutical company's—control mechanism. Quality can be assured by supplier certification, audits, and periodic reviews. Using XaaS suppliers that have established life sciences business and relevant compliance practices provides a significant advantage. Not only does this transfer most of the compliance work, but it presents less potential quality risk.

The ISPE GAMP® guidelines on risk-assessment analysis for computerized systems address some of these topics.[10] The World Health Organization[11] also addresses cloud-based service provider quality agreements. Britain's MHRA[12] was the first agency to dedicate an entire guidance section on managing cloud-based services. All of these organizations recommend using suppliers with specific life science certifications and governing them via an SLA. Naturally, it is also important that the pharmaceutical companies have a solid program in place to confirm compliance.

Cybersecurity, another point of major concern, often makes companies reluctant to trust cloud-based systems. Cloud technologies have gone through focused development, however, and are now considered more secure than on-premises infrastructure.[13] They have also adopted the latest security technologies and employ legions of security experts for continuous improvement. Many of these advances have been driven by industries that take data security and privacy very seriously, such as the finance and health care sectors, when they adopted cloud-based systems.

DATA INTEGRITY IN A CLOUD XaaS MODEL

Just as the traditional qualification process must be rethought in a cloud-based XaaS model, so must the approach to data integrity. This requires a holistic quality approach based on ALCOA and GAMP principles.

ALCOA—an abbreviation for the data properties *attributable, legible, contemporaneous, original,* and *accurate*—is used in regulated industries as a framework to ensure data integrity. ALCOA is vital for good documentation practices (GDP), and should be considered in any cloud-based solution.

Attributable: Secured protocols and certification keys are pillars of cloud technologies. Encryption algorithms (pairs of keys, private and public keys, or the recent blockchain technologies) are designed to ensure the authenticity of the message source. Protocols such as hypertext transfer protocol secure (HTTPS) or message queuing telemetry transport (MQTT) further facilitate the attribution task by using security certificates or peer-to-peer architectures.

Legible: Big data is built on the structures that encapsulate the data. *Legibility* means that these structures are easy to interpret. A common implementation is JavaScript object notation (JSON), which organizes data in plain text based on pairs of keys and values. Because these structures include definitions of data types, they do not need additional markup as is common in XML. JSON offers a leaner interpretation mechanism that makes legibility easier and uses less computational power.

Contemporaneous: Typical manufacturing process data is collected in real time and in frequencies measured in milliseconds. Many instruments are used to ensure the high availability, precision, synchronization, and time localization of the measurements. Network time protocol, local configurations, and fast internet infrastructures ensure that the measurement is in the right "envelope" and marked correctly. Raw data can be recorded instantaneously and sent to the cloud stamped with the time at which it was collected. If connectivity is lost, communication protocols such as MQTT can delay storage while maintaining the original time stamp.

Original: Several electronic tools can verify data originality data, using encrypted certificates. Original records can be wrapped with specific metadata (e.g., unit of measure, regional settings, geolocation) that provide context. Data-transfer models such as peer-to-peer or socket communications can guarantee that data is original.

Accurate: Logic rules can check data quality at the point of capture and before storage in the data lake. The challenge is to control changes to the data, which is not available in all big data environments. Change management is critical, and features built into big data solutions should include full audit trail, e-signatures, and security.

When applying the ALCOA principles and defining risk profiles in a big data environment, it's important to understand the data flow and corresponding decision flow that are at the core of advanced analytics in this environment (Figure 2).

To support big data solutions for regulated environments, ALCOA principles should be enhanced with the additional principles *complete, consistent, enduring,* and *available*, a combination known as ALCOA+.

Completeness and **consistency** should be implemented with data definition schemas and verification protocols as part of the system design.

Enduring: Cloud platforms with automatic backup and data replication across geographies practically guarantee "eternal perdurability." Data persistence is guaranteed as part of the data management system; each byte is triplicated in three different locations and on three different data storage systems.

Available: Data is always available within the defined retention periods, with no need for backups or data management.

Traditional checklists and assessments are not practical in a cloud-based services environment, since data technologies provide inherent mechanisms that comply with ALCOA+ principles by design (encryption, security, obfuscation, surveilled storage, access rules and rights, etc.), as represented in Table B. These systems can be qualified for a GMP environment in which data integrity is an inherent feature of the architecture. Every action and change in this self-monitoring system is stored in the GMP data lake. And here's the breakthrough promise: By applying AI algorithms to the data, the system can detect compliance levels and trigger an alert about any nonconformities. It is the same principle as continuous process validation, simply applied to the data itself.

IMPLEMENTING A CLOUD AND BIG DATA APPLICATION

Pharma and biopharma companies considering the switch to big data and cloud-based systems may be challenged by their lack of experience and best practices for these new technologies. They may also lack expertise with some of the more disruptive knowledge management features, such as AI. As a result, new strategies for qualifying and maintaining GMP and GDP have to be adopted:

- ☐ Start small to establish confidence with system quality and compliance. This is new technology, and probably unknown to most of the manufacturing organization. That makes the adoption and learning curve steeper.

TRANSFORMING CURRENT PHARMACEUTICAL MANUFACTURING TO PHARMA 4.0 REQUIRES A NEW APPROACH TO MANUFACTURING AND PROCESS DATA CAPTURE

It's critical that key stakeholders buy into the value that the new solution brings. A good practice is to start with low-friction deployments where the GMP impact is minimal. Applications for preventive maintenance, environmental monitoring, or energy savings are good candidates.[1] Other options are tasks that require repeatability and scalability. The FDA, for example, transformed report documents into digital data in a cloud-based solution with 99.7% accuracy, reducing costs from $29 to $0.25 per page.[14]

□ Establish a close and effective working relationship with the SaaS supplier. Because this is new technology for many organizations, it is imperative that all parties work together with a common goal in mind. The supplier must be fully transparent about the maturity, strengths, and weaknesses of the offered solution, understand the pharmaceutical business, and facilitate the tasks needed to qualify the infrastructure, platform, or service, as well as validate the processes. An SLA that supports a good working relationship with the SaaS supplier is important. While compliance remains a requirement, the implementation responsibility will shift to the provider.

□ For services related to artificial intelligence algorithms, indexing, or aggregation processes that require specific data calculations, the provider should be ready to share the verification process used to qualify data operations. It may be valuable to consider new innovative methods for system qualification and the use of the ALCOA+ principles.

□ A risk-based approach is a valuable tool when designing a big data, cloud-based solution. A good working relationship between the supplier and the organization is necessary to build trust; this is especially true for the QA organization.

□ AI, ML, and DL tools must be considered standard for analytics with very large data sets, some of which have already been recognized by the European Pharmacopoeia. The results of these analytical techniques are models that can be validated with the same rules applied to traditional statistics.

□ Make sure to utilize the automation and data integrity tools built into the cloud and big data infrastructure.

INDUSTRY ADOPTION

Cloud-based, big data analytics technologies are rapidly gaining acceptance in the biotech, pharmaceutical, and medical device industries. Industry 4.0 technologies have given rise to the Pharma 4.0 initiative led by ISPE. The promise of this new paradigm is consistent quality by use of a practical science-based approach. Predictions for 2020 and beyond[8] describe the following scenarios for life sciences and manufacturing activities:

By 2020:
□ Of the current AI proofs of concept in the top 50 life science R&D organizations, 20% will move into production.
□ A major technology provider will provide independent digital health services directly to patients.
□ Business function–led ML initiatives will be at least twice as successful as enterprise IT–led projects.

By 2021:
□ Three-quarters of the top life science R&D IT organizations will create data connections and business partnerships with nontraditional ecosystems.

By 2022:
□ Precompetitive, shared, and open research IT platforms will support at least 25% of life science distributed innovation environments to connect to an ever-growing R&D ecosystem.

XaaS ALTERNATIVES PROVIDE NEW METHODS THAT CAN MEET INDUSTRY REQUIREMENTS FOR LARGE DATA STORAGE AND COMPUTING RESOURCES AT A REASONABLE PRICE

Adoption is progressing at a rapid pace. There is a new energy in the industry based on data, not applications. These new technologies are not only easier to qualify, but can solve many of the industry's current data integrity challenges. Figure 3 shows that FDA warning letters, both inside and outside of the United States, increasingly include references to data management and data integrity.[15]

CONCLUSIONS

As always, progress is unstoppable. Industry 4.0 technologies will facilitate a generational change in manufacturing. Digitalization with big data, AI, cloud computing, continuous connectivity, advanced analytics, and the IIoT is here to stay, and the digital manufacturing plant is becoming a reality. In the more conservative biotech and pharmaceutical industries, adoption of these breakthrough technologies is being tested cautiously. As this happens, the Pharma 4.0 concept will gain momentum, especially for the practical application of GMP within this new reality. The cloud, big data, and AI make data an asset; this offers great potential for process optimization, science-based approaches, continuous process verification, and more. The ability to move from reactive to predictive, from strict to adaptive process control creates "smart" manufacturing in which advanced information and manufacturing technologies add flexibility to physical processes.[16]

Current data integrity best practices and guidelines are applicable in the XaaS big data environment, where inherent data management technologies automatically monitor compliance. AI technologies that analyze manufacturing data also allow constant monitoring and risk analysis using advanced analytics to find and segregate suspect data. Continuous data capture gathers data from any manufacturing or production source with no overall structure or context. Neural networks and pattern-recognition algorithms are used to train the system about right or wrong scenarios. AI algorithms discover context and patterns to detect nonconforming behaviors, unexpected errors, and inconsistencies. Examples of these solutions have already been developed in the predictive maintenance field with successful results.[17] The same predictive techniques are now being applied to detect and prevent data integrity issues.

The Pharma 4.0 paradigm has given manufacturing intelligence a new meaning. We can now truly achieve predictive and adaptive systems that manage the manufacturing process based on science. The future is here; climbing the steep part of the adoption curve should be our focus. ‹›

References

1. Herwig C., C. Wölbeling, and T. Zimmer. "A Holistic Approach to Production Control: From Industry 4.0 to Pharma 4.0." *Pharmaceutical Engineering* 37, no. 3 (May-June 2017).
2. US Food and Drug Administration. Guidance for Industry. Draft Guidance. "Data Integrity and Compliance with cGMP." April 2016. https://www.fda.gov/downloads/drugs/guidances/ucm495891.pdf
3. European Pharmacopoeia 9.0. Chapter 5.21. "Chemometric Methods Applied to Analytical Data." 2017.
4. US Code of Federal Regulations. Title 21, Chapter I, Subchapter A, Part 11: Electronic Records; Electronic Signatures. 4 April 2017. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=11&showFR=1
5. US Code of Federal Regulations. Title 21, Chapter 1, Subchapter C, Part 210: Current Good Manufacturing Practice in Manufacturing, Processing, Packing, or Holding of Drugs; General. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=210
6. US Code of Federal Regulations. Title 21, Chapter 1, Subchapter C, Part 211: Current Good Manufacturing Practice for Finished Pharmaceuticals. https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfcfr/CFRSearch.cfm?CFRPart=211
7. European Commission. Health and Consumers Directorate-General. EudraLex, Volume 4, Annex 11. "Computerised Systems." 30 June 2011. https://ec.europa.eu/health//sites/health/files/files/eudralex/vol-4/annex11_01-2011_en.pdf
8. Laney, Douglas, and J. Ankush. "100 Data and Analytics Predictions through 2021." Gartner. 20 June 2017. https://www.gartner.com/events-na/data-analytics/wp-content/uploads/sites/5/2017/10/Data-and-Analytics-Predictions.pdf
9. Kayyali, B., D. Knott, and S. Van Kuiken. "The Big Data Revolution in US Health Care: Accelerating Value and Innovation." McKinsey & Company. April 2013. https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/the-big-data-revolution-in-us-health-care
10. International Society for Pharmaceutical Engineering. GAMP Guide: Records & Data Integrity. March 2017. www.ispe.org
11. World Health Organization. "Annex 5: Guidance on Good Data and Record Management Practices." WHO Technical Report Series no. 996. 2016. http://apps.who.int/medicinedocs/documents/s22402en/s22402en.pdf
12. UK Medicines & Healthcare Products Regulatory Agency." "'GXP' Data Integrity Guidance and Definitions." Revision 1. March 2018. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/687246/MHRA_GxP_data_integrity_guide_March_edited_Final.pdf
13. Google Cloud. "Survey Report: Behind the Growing Confidence in Cloud Security." September 2017. https://lp.google-mkto.com/rs/248-TPC-286/images/SMR%20Google%20Cloud%20Security%20Survey%20Report,%20Final%20Approved%20Version,%209.8.17.pdf
14. Amazon.com. "AWS Cast Study: US Food and Drug Administration (FDA)." https://aws.amazon.com/solutions/case-studies/us-food-and-drug-administration/?nc1=h_ls
15. US Food and Drug Administration. "Warning Letters." https://www.fda.gov/ICECI/EnforcementActions/WarningLetters/default.htm
16. Davis J., et al. "Smart Manufacturing, Manufacturing Intelligence and Demand-Dynamic Performance." *Computers & Chemical Engineering* 47 (20 December 2012): 145–156.
17. Mobley, R. K. *An Introduction to Predictive Maintenance.* 2nd edition. Elsevier, 2002.

About the authors

Toni Manzano, PhD, is R&D Director and Co-Founder of bigfinite, inc., which provides big data and AI SaaS platform for the biotech and pharma industries. Since 1996 he has led software projects for international pharmaceutical companies covering the entire supply chain and production process. He has also led projects implementing solutions based on ICH Q8, Q9, and Q10, executing audit, qualification, and validation tasks. A physicist, he has worked as a researcher at the University of Barcelona, where he also teaches big data and AI for life sciences students. He has written numerous articles in the pharma field and holds a dozen international patents related to the encryption, transmission, storage, and processing of large volumes of data for regulated environments in the cloud. He has holds a master's degree in information and knowledge society, and has done post-graduate work in quality systems for manufacturing and research pharmaceutical processes. He has been an ISPE member since 2015.

Gilad Langer, PhD, is an accomplished business leader with 25 years of experience in managing engineering services, technical operations, sales, business development, and marketing. He has deep domain expertise in manufacturing information systems for the life sciences industries, where he has successfully developed engineering services and consulting businesses. His focus and passion is engineering and delivery of complex automation and manufacturing software solutions, most recently cloud-based manufacturing intelligence SaaS solution using big data and AI. He has also served as a trusted adviser and business consultant in the areas of technology directions, industry strategy, and software implementations. He has honed a systematic and strategic approach throughout his career in roles that included consulting, sales, marketing, product strategy, project management, software development and implementation, and research. He has been an ISPE member since 2010.