Journal of Pharmaceutical Sciences 109 (2020) 1427-1438

Contents lists available at ScienceDirect

Journal of Pharmaceutical Sciences

journal homepage: www.jpharmsci.org

Review

Transitioning Chemistry, Manufacturing, and Controls Content With a Structured Data Management Solution: Streamlining Regulatory Submissions

Marquerita Algorri [1, 2], Nina S. Cauchon [1, *], Michael J. Abernathy [1]

[1] Department of Global Regulatory Affairs and Safety—CMC, Amgen Inc., Thousand Oaks, California 91320
[2] University of Southern California, School of Pharmacy, Los Angeles, California 90089

ARTICLE INFO

Article history:
Received 26 July 2019
Revised 3 January 2020
Accepted 23 January 2020
Available online 29 January 2020

Keywords:
regulatory science
machine learning
artificial intelligence
database(s)
automation

ABSTRACT

The process of assembling regulatory documents for submission to multiple global health agencies can present a repetitive cycle of authoring, editing, and data verification, which increases in complexity as changes are made for approved products, particularly from a chemistry, manufacturing, and controls (CMC) perspective. Currently, pharmaceutical companies rely on a workflow that involves manual CMC change management across documents. Similarly, when regulators review submissions, they provide feedback and insight into regulatory decision making in a narrative format. As accelerated review pathways are increasingly used and pressure mounts to bring products to market quickly, innovative solutions for assembling, distributing, and reviewing regulatory information are being considered. Structured content management (SCM) solutions, in which data are collated into centrally organized content blocks for use across different documents, may aid in the efficient processing of data and create opportunities for automation and machine learning in its interpretation. The US Food and Drug Administration (FDA) has recently created initiatives that encourage application of SCM for CMC data, though many challenges could impede their success and efficiency. The goal is for industry and health authorities to collaborate in the development of SCM for CMC applications, to potentially streamline compilation of quality data in regulatory submissions.

Introduction

Modern society is becoming increasingly reliant upon the application of “big data” analytics, automation, and artificial intelligence (AI) as resources for augmenting human knowledge and revolutionizing product design. Accordingly, as other industries have continually transformed and modernized their products, the complex processes that comprise pharmaceutical

Abbreviations used: AI, artificial intelligence; ANDA, abbreviated new drug application; BLA, biologics license application; CBE30, change being effected in 30 days; CCMS, component content management system; CPT, common protocol template; CQA, critical quality attributes; CSR, clinical study report; CTD, common technical document; DITA, darwin information typing architecture; EC, established conditions; eCTD, electronic common technical document; EMA, European Medicines Agency; FDA, US Food and Drug Administration; FHIR, fast healthcare interoperability resources; GMP, good manufacturing practices; HL7, health level 7; ICH, International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use; IFU, instructions for use; IoT, internet of things; IT, information technology; KASA, knowledge-aided assessment and structured application; MELLODDY, machine learning ledger orchestration for drug discovery; MA, marketing applications; NDA, new drug application; NIH, US National Institutes of Health; NOC/c, notice of compliance with conditions; OPQ, Office of Pharmaceutical Quality; PACMP, post-approval change management protocols; PDUFA, prescription drug user fee act; PLCM, product lifecycle management; PMDA, pharmaceuticals and medical devices agency; POC, proof of concept; PQ/CMC, product quality/chemistry, manufacturing, and controls; PQRI, product quality research institute; PQS, product quality system; PRIME, priority medicines; QOS, quality overall summary; SAP, statistical analysis plan; XML, eXtensible markup language.

Conflicts of interest: All authors are employees of Amgen Inc. (except M.A. who is a summer intern). All authors contributed to the writing of the report and approved the final version of the manuscript and the decision to submit the manuscript for publication.

* *Correspondence to:* Nina S. Cauchon (Telephone: +1 805-447-1569).
E-mail address: ncauchon@amgen.com (N.S. Cauchon).

https://doi.org/10.1016/j.xphs.2020.01.020

development and regulation must also evolve to fit the changing needs and demands of a data-rich, global environment. In comparison to the automotive, aerospace, computer, and consumer packaged goods industries, the pharmaceutical industry has historically had low investment in design of novel equipment, improvement in annual productivity, first-pass yield, and value-added time for labor, while struggling with long production times.[1] The manufacturing sector has been traditionally slow to change due in part to rigid regulatory requirements that can encumber the development of advanced technologies, as innovative approaches carry inherent and potentially unforeseen risks to product quality. Similarly, manufacturing equipment has limited adaptability to change and is often designed for a distinct, predesignated purpose. Although there has been notable innovation in the use of AI, machine learning, and automation to improve the efficiency of data analysis and management during the drug discovery process, there has not been as much progress in automating the collection, organization, storage, and analysis of pharmaceutical manufacturing and quality data used for regulatory submissions.[2] Introduction of new methodologies for storing, analyzing, authoring, and reviewing data may aid in the introduction of new technologies, as machine learning and AI are most easily applied when data are in a "structured" format. Ideally, changes in how information is stored, analyzed, authored, and reviewed should be enacted collaboratively between global regulatory agencies and industry.

In this review, we will specifically examine structured content management (SCM) systems as potential solutions for enabling mobilization of data, enhancing content reusability, reducing data integrity concerns, and providing opportunities to integrate automation and machine learning technologies. Though few significant advances toward using SCM for CMC data management have been publicly disclosed to date, SCM has been previously applied to support other functional areas within the pharmaceutical industry, such as clinical data management, supply chain management, and labeling. SCM has the potential to address a variety of current concerns in CMC data management. We will also discuss a number of challenges associated with its development and implementation that will need to be overcome to realize its potential.

Overview of Regulatory Challenges Faced Within the Biopharmaceutical Industry

Many of the methods and assays currently used during pharmaceutical manufacturing to evaluate product quality generate an abundance of data, as in-process testing as well as specification and stability testing parameters must be closely monitored and data recorded throughout the process. Depending on the company, a variety of in-house and commercially available information systems are used to capture these data. For some companies, much of this in-process, validation, and batch release data are destined for a data warehouse or a "data lake," a relatively new system for pooling structured and unstructured data directly from its source for storage on a cloud-based server.[3] Although "data lakes" represent a significant improvement in data storage by creating a single, unified repository, not all of the raw data are easily accessible, and the structured data that are accessible may still require manual querying.

Once collected and validated, the data generated by all manufacturing processes as well as quality assessments must subsequently be condensed and compiled into regulatory submissions to health authorities around the globe to obtain or maintain marketing approval for commercialization of human therapeutics or to initiate clinical trials with experimental therapeutics (Fig. 1). From a regulatory perspective, much of these data are destined for the Common Technical Document (CTD), an internationally recognized format created by the International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) for submitting documents for regulatory review. Use of the CTD is increasingly becoming a requirement in major markets, and in some regions, the CTD must be submitted electronically through an online portal.[3-5] This format is referred to as the electronic CTD (eCTD), and while the first version of eCTD was developed in 2002, eCTD did not become mandatory in select regions until 2016. As of 2019, many agencies are still accepting paper submissions, with some regions predominantly utilizing paper submissions. Both the CTD and eCTD separate product regulatory information into 5 modules with predetermined, numbered, and itemized sections and subsections. Module 3 displays all quality and chemistry, manufacturing, and controls (CMC) data, pertaining to product manufacturing, analytical methods, process development, specification testing, and stability of drug substance and drug product. This information is then summarized in the Quality Overall Summary (QOS), which appears in module 2.

The aforementioned "data lakes" where validated experimental data are stored can be designed to make data organization and input into eCTD more effective. A critical design factor that ought to be considered in the development of a "data lake" is building meta-data and associating the raw data through semantic relationships via related ontologies and data models. What semantically enabled data allow is significantly increased ease of accessibility and task automation. In other words, finding the data and information one is looking for is made easier and more effective. Another crucial enabling aspect of the semantically enabled data is to insert hooks, tags, and labels to mark the necessary elements (such as raw data tables, specifications, and graphs) that will end up in eCTD ahead of time, allowing the elements of eCTD to be automated or queried more efficiently. Although data lakes can be improved via internal data modeling, there are also consortia specifically focusing on publishing data standards that can be leveraged across biopharmaceutical companies. The Allotrope Foundation, Pistoia Alliance, and QUDT.org are among the key organizations that have membership or utility across companies that publish agreed upon data standards regularly for companies to adopt.[6-8] Although these technologies are still emerging, they carry significant potential for automating data management, organization, and analysis.

Even with assistance from a "data lake," the process of processing and entering data into the eCTD still requires substantial human oversight. Written narratives or descriptions are required along with justifications for specification criteria and shelf-life that are based on the data that have been collected and analyzed during drug substance and drug product development, scale-up, and clinical and commercial manufacturing. Once the required information has been written and appropriately formatted, a submission is made to a health authority for review. When submissions reach health authorities, agencies are required to collate and reassemble information submitted by the industry applicant, providing feedback in the form of data tables copied from the applicant's submission with accompanying text-based assessments authored by the reviewer. The submission and review processes, in their current format, require repetitious use of content through multiple cycles of retyping and reorganizing data, decreasing the workflow efficiency of both sponsors and health authorities, while increasing the chances of transcription errors. Every editing step requires substantive validation from the authors before publishing to ensure consistency across documents and data sets. Thus, the processes of maintaining data integrity and version control across documents

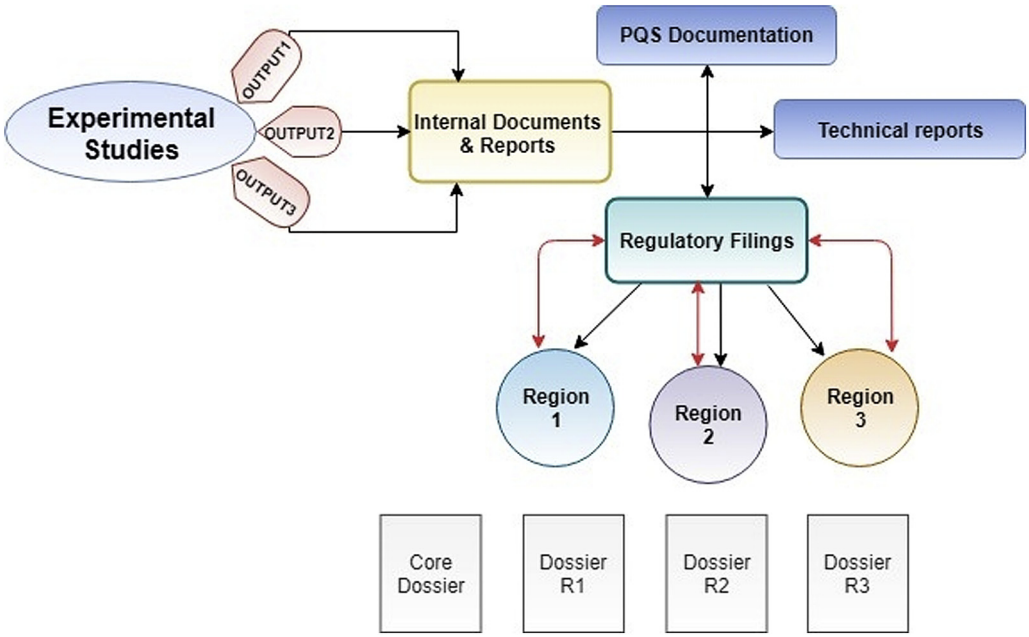

Figure 1. The current workflow for managing CMC regulatory submissions starts with output from experimental studies that is assembled into internal documents and reports. From these internal documents, the data are summarized and repurposed for product quality system (PQS) documentation, technical reports, and regulatory filings. Regulatory filings vary based upon region and are sent to global health authorities (black lines). The initial submission process begins a series of back-and-forth submissions between health agencies and industry (red lines), until the submission is approved. Once finalized, the documents can then be organized as a series of regionally variable core dossiers.

are intensive and time consuming for both the preparation and review of regulatory submissions.

Global Harmonization of Regulatory Requirements

If submitted marketing applications are approved for product commercialization, additional adjustments will likely need to be made to accommodate inevitable changes in manufacturing processes and facilities throughout a product's lifetime. From a CMC regulatory perspective, perhaps the most complex and labor-intensive aspect of managing regional regulatory submissions occurs after approval, once the product has entered the commercialization stage. When postapproval changes are made, the required reporting steps to health agencies vary by country. For drugs approved in the United States, when major changes are made, such as those that are highly likely to adversely affect the identity, strength, quality, purity, or potency of the product, a formal amendment must be filed and approved before the change is enacted.[9] While the United States, Canada, and EU have somewhat similar requirements for submission of major, moderate, or minor risk changes, Japan's regulatory body, the Pharmaceuticals and Medical Devices Agency (PMDA) has a different system that requires submissions for all changes, including low-risk changes.[10] More detailed comparisons of global postapproval change regulatory mechanisms in select major markets are presented in Table 1.

The need to submit modified versions of the same application to global regulatory agencies increases review burden for each individual health agency, in addition to increasing the workload of industry regulatory affairs personnel and other authors within the pharmaceutical industry. When assembling CMC submissions, many authors across departments and outside of regulatory affairs who can impart technical knowledge on drug substance, drug product, and manufacturing procedures may be required to contribute to authoring. These personnel may also be integrally involved in document change management and must re-review documents as they are revised and provide input even if changes are minimal. The process of building documents tailored to the varying requirements of each country is currently an intensive manual process, which is time consuming, requires substantial human involvement with little automation, and presents a risk of introducing errors.

Expedited Review Pathways

The lack of international regulatory harmonization can delay worldwide patient access to products that fulfill unmet medical needs. Accelerated regulatory pathways that grant expedited review and approval are of increasing relevance as global health authorities continue to establish new programs intended to reduce review times for these therapeutic agents.[12] In the United States, Priority Review was first introduced in 1992. Within the past 20 years, the FDA has since expanded the number of available accelerated designations to include Fast Track, Accelerated Approval, Breakthrough Therapy, and Regenerative Medicine Advanced Therapy. Globally, a variety of expedited review pathways are now available for impactful therapies for serious and rare diseases, including but not limited to the EMA's Accelerated Assessment of Priority Medicines (PRIME); Sakigake in Japan; Notice of Compliance with Conditions (NOC/c) in Canada; and Priority Review programs in Australia, India, Russia, and China. Accelerated pathway requirements often differ between regions and can create disparities in application materials between countries, though the CMC/Quality development aspects are frequently on the critical path to gaining marketing authorization approval.[13] There is overall less time during development to generate product quality data, including key aspects such as stability data, product characterization studies, scale-up, and process characterization and validation. As a result, reduced product development timelines require a larger number of postapproval variations, adding to the complexity of regulatory submission management in the lifecycle phase. Moreover, products that qualify for expedited review in one country may not meet requirements in another country, putting the product on a

Table 1
Summary of Regulatory Requirements in Major Markets for Postapproval Change Notification

Change-Associated Risks	US-FDA	EU-EMA	Japan-PMDA	Canada-HEALTHCANADA
High: Significant potential to impact product quality and safety	Major change: requires prior approval supplement (PAS) before changes can be made	Type II variation: requires application for approval of variation	Partial change: requires application for approval of variation	Level I—supplements: requires application for approval of variation
Medium: Moderate potential to impact product quality and safety	Moderate change: requires notification via Changes Being Effected in 30 Days (CBE30) Supplement; changes can be made 30 d after notifying FDA	Type IB variation: requires notification; changes can be made 30 d after notifying EMA	Minor change, 30-d notification: requires notification within 30 d of change	Level II—notifiable changes: requires notification and issuance of a no objection letter (NOL)
	Moderate change: requires Changes Being Effected (CBE) Supplement at time of change	Type IA_{IN} variation: requires immediate notification		
Low: Minimal to no potential to impact product safety and quality	Minor change: requires notification in annual report, no notification at time of change	Type IA variation: requires notification within 12 mo of change		Level III—annual notification: requires notification in annual report, no notification at time of change
				Level IV—record of changes: requires manufacturer or sponsor to maintain a record of changes, no submission required

Partially adapted from.[11]

different schedule with respect to manufacturing and validation data needs.

New Modalities

Drug developers have achieved significant progress toward innovating novel therapeutic modalities beyond traditional small-molecule synthetic compounds and well-characterized biologic products, as demonstrated by the first gene therapy approvals in the United States in 2017, approval of 9 oligonucleotide-based therapies as of 2019, the first allogenic stem cell therapy approval in Europe (Alofisel—Takeda), approval of oncolytic viral therapies (IMLYGIC® – Amgen), and a variety of other accomplishments in cell and gene therapy.[14] Many of these therapies are also examples of personalized medicines, which are designated for patient-specific disease biomarkers that may be heterogeneously expressed across patient population (e.g., HER2 or PD-L1 in cancer subsets; CYP2D6 in Huntington's disease and tardive dyskinesia). Alternatively, personalized medicines may be autologous cell products that are manufactured in individual lots by using the patient's own cells to target unique cancer cell surface markers.

As drug modalities advance beyond the scope of current knowledge and continue to incorporate personalized approaches, manufacturing processes will shift away from well-characterized approaches, adding additional uncertainties and complexities to risk assessment and mitigation strategies. A dynamic data storage solution and an accompanying machine learning algorithm that can utilize scientific concepts and historical product data to infer risk could be developed into a valuable toolkit for use by the manufacturer throughout the risk management assessment process. An information technology (IT) solution capable of organizing and distributing data on all manufacturing processes in a holistic manner could also be integrated within future data management systems, with the potential to inform industry on future process development for novel products. Innovation in data management is becoming increasingly necessary for accommodating an expanding landscape of complex data. More specifically, organizing information into reusable content blocks using an SCM-based organizational framework may assist industry in providing the necessary infrastructure for optimizing collection, storage, and submission of CMC data for regulatory applications by allowing for enhanced synchrony and establishing connections across documents.

SCM for Regulatory Documents

Current authoring and publication conventions across pharmaceutical companies and regulatory bodies require written documents, including data files, to undergo review, verification, and approval before distribution. The approved information, often required in multiple internal and external documents, must be manually imported into separate documents and then reverified and reapproved. This compiling process limits a company's ability to compare data collected for different products and submissions. For example, if a company is interested in viewing a regulatory submission for a legacy product in the process of the creation of a new submission, the data often have to be queried manually, typically by requiring personnel to physically find, view, and interpret the older documents to establish a frame of reference for creating a new document. It is similarly challenging for health authorities to provide consistent, objective feedback for products within the same modality or mechanism of action, as accessing sources documenting prior decision making and review documents is a laborious task. In addition, if information provided by the sponsor changes or requires updates or is sent to different regulators, it must be manually updated and reverified in each document in which it appears, as information is not linked across documents to enable real-time editing. As a result of this process, there are multiple versions of a document that must be maintained, increasing opportunities for error and distribution of outdated information.

SCM solutions offer opportunities for enhancing review and organization of documents by providing the framework for authoring content that is easy to adapt into multiple documents and standardize. SCM systems store information on a single server and group blocks of information into content libraries to allow for uniform importation of approved data across documents. If the imported data are edited, changes can be tracked electronically, allowing for the generation of a digital data trail, which provides insight into who revised the document and how it was changed.

The resulting edits can then be validated and automatically applied to all documents in which that content block appears, allowing for increased consistency and uniformity across documents. Typically, SCM systems allow for content that is both human and machine readable, increasing the potential flexibility for conducting downstream tasks. For example, data can still be browsed and analyzed manually if desired, but a crucial advantage of SCM is the ability to assess and compile large amounts of data automatically, in different configurations, as needed. The content management software company, DitaExchange, has fittingly compared SCM solutions to Legos®: SCM categorizes information into sets of connectable "bricks" that are capable of structural remodeling and repurposing based on specific needs.[11]

Many SCM systems are built around and extensively utilize a Web-based programming language and file format known as eXtensible Markup Language (XML). XML provides, in addition to body text or content, built-in meta-descriptions of the content within its framework, thereby conferring structure to a document or group of documents by labeling their parts, enhancing categorization and searchability. XML files offer innate flexibility for organizing different types of elements and manipulating data because the language contains few defined tags, or labels for different types of information, and allows users to develop their own tags.[15] In addition, XML is compatible with plain text files and can be edited using basic word-processing software, including Microsoft Word. The current eCTD application uses an XML-based backbone to organize PDF submission components.

Advanced SCM solutions are not the industry standard and have not been universally adopted across companies, as substantial resources are required for their development and implementation. If pursued, use of SCM creates a sizable change in the existing workflow and requires authors and IT-personnel to adapt to new methodologies and potentially unfamiliar technology, which can present logistical difficulties. Several biopharmaceutical companies are currently navigating these challenges by exploring XML-based SCM technologies to manage aspects of their data handling and document authoring operations. Notably, most prior forays into SCM utilization for data management have focused on clinical data or labeling, as opposed to CMC, quality, and manufacturing data. Several examples of SCM use in pharmaceutical companies will be explored in the following sections.

Select Examples of SCM for Pharmaceutical Labeling & Package Inserts

Sanofi, an international pharmaceutical company headquartered in France, has undertaken multiple endeavors to structure data using Darwin Information Typing Architecture (DITA), an XML-based SCM system created in 1999 by IBM that allows for information classification and grouping, mapping, and reuse of content and contains descriptive metadata to aid in structuring and presentation of data elements.[16] Sanofi partnered with the software company, DitaExchange, and the IT integration consulting firm, ArborSys, to engineer a structured approach for managing global postmarketing labeling data.[17] Labeling information is maintained as a core dossier that is then modified based on region-specific requirements. These documents are living documents that are subject to change based upon, for example, the emergence of a new serious adverse event or approval of a new indication or dose. The same "copy-paste" document management problem described previously also applies here, creating problematic inconsistencies between sequential versions of documents. ArborSys's SCM technology divided components of the labeling document into content blocks and created a branching model to demonstrate interactions between blocks. DitaExchange provided a Microsoft Word—accessible platform for writing and editing of content blocks. The results of this pilot project resulted in significant time savings due to more efficient document validation and publishing, as well as a reduction in time and effort needed to respond to health authority feedback.

In another example, Medtronic, an international medical device company, has used SCM to generate labeling-related documents to accompany medical devices. Medtronic partnered with Vasont Systems, a content management software company, to optimize document authoring across therapeutic areas and business units. Medtronic began using Vasont's XML-based component content management system (CCMS) as early as 2003 for neuromodulation device data and has subsequently expanded across several different departments from 2005 to 2013. Types of data managed include package inserts, instructions for use, device and software manuals, and patient guides. Owing to diversity in document formats and intended audiences for these documents, the CCMS needed to be sufficiently flexible. For example, although the user information for neuromodulation devices, such as deep brain stimulators, is targeted toward surgeons and medical professionals, devices managed by the diabetes group must include patient-accessible labeling as most of these devices are used by the patient. Different product needs may result in discrepancies between necessary sections in documents. In addition, the CCMS is compatible with multiple different word-processing tools to accommodate departmental preferences. Medtronic has also seen benefits in increasing language translation efficiency and consistency when managing documents using CCMS.[18]

Select Examples of SCM for Clinical Data Management

There are multiple ongoing efforts across companies to use SCM for different aspects of clinical data management to maximize data reusability and reduce authoring time. For example, Amgen, a global biopharmaceutical company headquartered in the United States, has been exploring SCM solutions since 2012 for management of clinical data. Most recently, in 2018, Amgen partnered with ArborSys and Greenwich Biosciences to develop a proof-of-concept (POC) framework for authoring and organizing clinical protocol, statistical analysis plans, and clinical study reports.[19] Before implementation, Amgen had been using SMART templates, which provide an XML backbone and allow metadata tagging of content to aid facilitate reusability within documents and manual export of approved content for insertion into separate documents. However, the content in SMART templates does not connect directly to a central database for interdocument connectivity, which is a significant limitation. The POC SCM system expanded upon the capability of SMART templates by allowing for more connectivity through storage of content blocks within a central database, creating opportunities for content reuse without requiring authors to manually export and import content sections. The POC study met the key functional requirements of conferring reusability, enabling tracking of content origin and changes, maintaining prepopulated document templates, and establishing searchable content libraries. The success of the POC study leads to the recommendation of a larger production pilot study before future implementation of SCM for clinical data management.

As discussed previously, Sanofi has collaborated with experts at ArborSys and DitaExchange to implement SCM solutions for a variety of functions, including clinical data management. The partnering organizations collaboratively developed a prototype tool with an accessible user interface for more efficient authoring of patient narratives, a component included in clinical trial reports that provides details on individual patients who experienced serious adverse events, adverse events of interest, or

discontinuation in the study.[11] The resulting tool was constructed using a SharePoint-based platform that allowed authoring in Microsoft Word and included a plug-in to enable conversion of text into XML format, permitting the DitaExchange software to integrate content into its management system, which incorporated existing features of SharePoint.[20] The overall results of the pilot project indicated that use of a SCM system saved time, reduced costs, and allowed for reuse of information for other documentation. Notably, authoring time for a single patient narrative was reduced to 30 min versus 6 h before SCM integration and 65% of patient narrative content was reusable for other documents.[11]

Because many of the previously discussed data management challenges are common across industry, collaborative efforts between companies for generating standardized, structured documents have also emerged with relative success. TransCelerate BioPharma, a nonprofit organization with representation from 20 multinational companies, has established several key initiatives toward data standardization for regulatory submissions and mobilization of data from cloud-based servers, focusing primarily on clinical trials data and associated documents.[19] Among TransCelerate's key accomplishments is the creation and implementation of the Common Protocol Template (CPT), intended to harmonize and standardize clinical trial protocols. Importantly, the CPT provides structure for clinical trials protocol through a Microsoft Word—based authoring format with an XML-backbone template containing standardized headings and accompanying defined variables. In addition to 47 built-in variables, the CPT allows users to create unique variables to meet the specific needs of the user. The content generated within the CPT is reusable across documents and can be used to generate registries on Clinicaltrials.gov. CPT was created in collaboration with the National Institutes of Health (NIH) and FDA, which have formally endorsed the CPT eProtocol tool for NIH-funded phase 2 and phase 3 IND studies.[20]

Select Examples of SCM for CMC Data Management

In comparison to the numerous examples of SCM use for clinical and labeling data, there are relatively few publicly disclosed instances of SCM application for management of CMC data. Accordingly, there is substantial unrealized potential for CMC data adaptation into SCM format, and particularly, a system that allows for programmable elements that are human and machine readable for data input and analysis of quantitative quality specifications.

Although supply chain management is a distinct function and not considered to be an integral aspect of CMC regulatory affairs, it is a factor related to product manufacturing for which SCM solutions have been used. For example, Amgen is currently utilizing the MarkLogic database platform, which provides an SCM framework, to manage supplier and product information, organize carrier routes and information, and track temperature control data and methods.[21] Using MarkLogic, Amgen has been able to connect data located in disparate documents to observe a more complete model of product shipping and batch distribution, allowing the company to evaluate patterns within their supply and distribution methods.

From a regulatory perspective, the FDA, EMA, and ICH have also made strides toward conceptualizing standardized systems that incorporate the use of SCM for management of CMC components of regulatory dossiers. However, many of these changes are in their ideological infancy and lack formal structure, including IT platforms for mobilizing efforts toward standardization. The lack of a current, defined framework confers significant ambiguity toward the ease of applicability and implementation of these proposed programs on behalf of industry. Although SCM represents a potentially revolutionary method for collecting and analyzing data, global harmonization of SCM strategies utilized across regulatory submissions is paramount to leveraging its success. The following sections will focus primarily on new FDA initiatives in the United States that focus on developing SCM solutions for submission and review of regulatory applications.

The FDA's KASA Initiative

In the United States, the FDA is taking the initial steps to integrate an SCM-based approach for data collection and review of CTD Module 3 quality and CMC data. The agency is piloting several initiatives that will facilitate a shift away from the unstructured narrative summary format for review of regulatory submissions and amendments in favor of quantitative, data-rich entries that are machine readable and human accessible. Among the most concrete actions that have been taken toward this forward-reaching goal is the creation of FDA's Knowledge-Aided Assessment and Structured Application (KASA) initiative, announced in June 2018 by former FDA Commissioner Scott Gottlieb and spearheaded by the Office of Pharmaceutical Quality (OPQ).[22,23] KASA was created in response to increasing numbers of Abbreviated New Drug Application (ANDA) submissions and mounting public pressure to increase availability of affordable medications, as the presence of generic medications on the market increases competition and decreases costs for consumers. Additional details were provided in a public FDA meeting in September 2018 and at the 2019 Fourth Annual Product Quality Research Institute (PQRI)/FDA Conference.[24,25]

In addition to consumer benefits, there are elements within KASA that aim to reduce the workloads of health agency regulators and submission authors in industry. While the backlog of generic applications awaiting review has markedly improved, decreasing from 1154 unreviewed ANDAs in 2012 to 148 unreviewed ANDAs in 2018, the agency continually aims to improve and increase its efficiency.[26,27] The FDA has indicated that the structured application system proposed by the KASA initiative will help sponsors submitting ANDA applications for small-molecule drugs to provide high-quality, complete data to avoid multiple review cycles that create lengthy delays in the commercialization approval process. A more structured application will reduce ambiguities in regulatory expectations and increase the likelihood of approval during the initial review cycle.[24] Conversely, KASA's application is associated with a variety of critical challenges that adversely affect its success and productivity, which will be discussed in greater detail in the Section SCM Implementation Challenges.

KASA and the Future of Module 3 eCTD Data

The current review process for CMC data relies heavily upon the expertise and experience of the agency reviewer in interpreting the data presented in the submission. Although historical data can be a powerful tool for determining risk and evaluating regulatory strategies, it is difficult to access, as the format of submissions and reviews complicates the act of comparing across products, companies, and regions.[28] Conducting risk assessment can also present significant challenges, as summaries are lengthy and can fail to reach a concise and well-defined conclusion.[24,28] KASA attempts to automate and standardize risk assessment by mobilizing raw data, automating risk assessment based on an algorithm, and providing prefilled content selection tools for the reviewer to recommend risk mitigation strategies. The approach is represented symbolically through a “house” schematic, where 3 distinct “pillars” of risk assessment rest on a “knowledge base” foundation and support an automated and standardized “knowledge-aided assessment” roof (Fig. 2).

Although the FDA has established a theoretical model for its KASA initiative, it has not yet provided an IT framework for industry

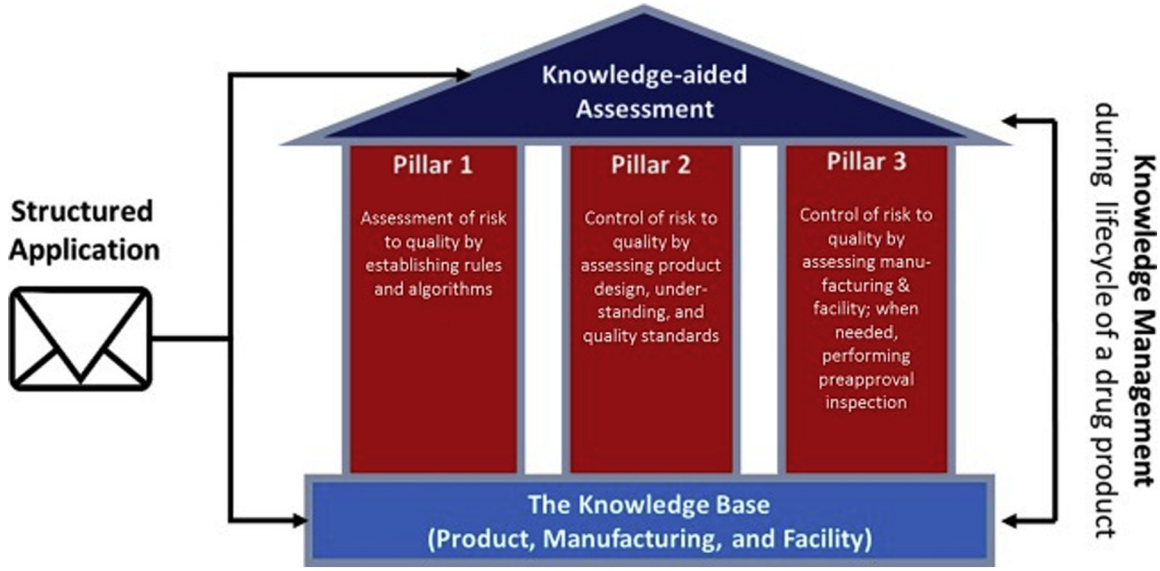

Figure 2. The Knowledge-Aided Assessment and Structured Application (KASA) schematic describes the 3-pillared approach. FDA has conceptualized to innovate and standardize its ANDA review process for small molecule generic drugs. Adapted from Yu et al.[28]

stakeholders to interface with, raising numerous questions regarding what KASA's implementation will look like, and how established manufacturing data reporting systems will need to change. Though the FDA is aiming to move toward automated risk assessment during review, industry must remain primarily in control of decision making regarding product-specific risk management. The Agency's plans for integration of KASA with previously announced FDA initiatives, such as PQ/CMC data standards, the QOS, as well as international guidelines, such as ICH Q12, are ambiguous, but there is potential for consensus between these regulations, as well as industry considerations, if proposed regulations are properly managed.[29,30]

Lifecycle Management

While KASA is intended to impact US ANDA applications, with probable impacts upon novel entity submissions in the future, the successful implementation of FDA's KASA will require significant global collaboration. Because the US-based KASA system will present a departure from the current workflow for interfacing with regulatory agencies, in an ideal situation, other regions would adopt similar systems. If this global element is neglected, industry will be required to submit fundamentally different applications in the United States versus the rest of the world, which may delay approval. As a potential partial remedy, the recently endorsed harmonized guideline ICH Q12: Technical and Regulatory Considerations for Pharmaceutical Product Lifecycle Management presents opportunities for incorporating key elements of KASA into submissions across regions. Specifically, ICH Q12 provides outlined strategies for implementing harmonized regulations for postapproval change management throughout a given product's lifecycle.[31] It establishes a variety of related documents to be included with the original MA that describe planned changes to the manufacturing process, thereby establishing a prearranged agreement between industry and health authorities that specific changes will be implemented and simplifying the review process for predicted postapproval changes. In addition, ICH Q12 suggests global harmonization of risk classifications (high, moderate, low) across countries for postapproval change requests that require prior approval or notification, or those that can be managed within the pharmaceutical quality system. To the same effect, the guideline introduces the concept of established conditions (ECs), which seeks to define critical process parameters that impact quality and, if modified, would require postapproval amendments or notification.

ICH Q12 also provides guidelines for Post-Approval Change Management Protocols (PACMPs), which are written protocols with validated acceptance criteria describing future changes to ECs that pharmaceutical companies intend to implement during the commercialization phase. The PACMP represents a structured approach to manage EC changes and can span multiple changes for multiple products, or single changes for one product, which can be submitted at the time of the MA or any time thereafter. Both ECs and PACMPs are integral components of the product lifecycle management (PLCM) plan, which intends to provide a centralized platform for predicting and informing regulatory authorities about planned changes to manufacturing processes for products in the commercialization phase. The PLCM is a living document that will require modification if postapproval CMC change submissions are filed. The proposed harmonized approach to managing manufacturing and analytical procedure changes will encourage unification of postapproval submission requirements across regions, enabling more effective change management solutions to provide faster access to higher-quality products. Integration of an SCM solution (such as discussed in this article) to track EC changes and PACMP modifications within the PLCM would assist in optimizing the global maintenance of this continually changing document.

PQ/CMC Standardization and Implementation of QOS

In the United States, the FDA indicated their interest in increasing standardization in regulatory submissions by releasing draft documents on standardizing pharmaceutical quality/chemistry manufacturing controls (PQ/CMC) data.[29,32] While KASA intends to change how submissions are evaluated by the FDA, the PQ/CMC initiative aims to transform how industry presents CMC data in submissions by harnessing a structured data approach. A draft document describing the FDA's plans to standardize select PQ/CMC data elements was published for public comment in the Federal Register on July 11, 2017.[32] This program would require module 3

data to be submitted in structured data elements, rather than in narrative and data table format in multiple PDFs, as the current eCTD format necessitates.

The draft documents released in 2017 identified specific CMC data elements and proposed formatting for future standardized submissions. For example, standards for specifications, such as the name, version, version date, and approval status date of the specification, and any additional comments, would be provided in text format, whereas the specification type (i.e., drug substance, drug product, raw material) and approval status would be provided in code. Although there is not much available information to date regarding the actual infrastructure required to implement the PQ/CMC system, as draft guidance documents are not targeted for release until 2020, the FDA has begun a proof-of-concept PQ/CMC project using Health Level Seven International's (HL7) Fast Healthcare Interoperability Resources (FHIR) that may provide preliminary insight on how the PQ/CMC data entry system could function, as described in the following.

FHIR & PQ/CMC

FHIR is an international standards framework that was constructed using a Web-based architecture and is based upon structured blocks of information known as "Resources." It was created to facilitate sharing of information across different health care entities, such as hospitals, clinical laboratories, physician's offices, and radiology laboratories, to provide more complete information about patients and enable searchability and filtering capability for electronic medical records. FHIR is compatible with XML files and JSON file formats. In its original context as a tool for clinicians, patient personal and clinical data are more likely to be displayed in JSON files, whereas laboratory results may be stored in XML files, creating additional steps for health care providers to compile patient data. FHIR provides a standardized format for viewing both file types.

The resources that comprise the FHIR framework are both human and machine readable and contain common definitions and metadata to provide descriptions, structure, and context.[33] These standardized Resource blocks (comprising categories such as "patients" and "medications") cannot be directly altered but can be modified through extension content blocks that can be created by users based on their specific needs. Resources and their extensions can refer and link back to one another and all content will be accessible across the FHIR network. The FDA is piloting a project using FHIR to collect and exchange PQ/CMC data from industry to regulatory agencies.[29,34] Although the project is still in POC stages, if it is implemented as a part of future FDA submissions for module 3 data, this will allow for potential direct linkages to other sectors of the health care industry and facilitate submission of data that is both machine and human readable, augmenting potential opportunities for automation.

As published on the software development platform Github in June 2019, the currently identified Resources being utilized in the initial draft of the PQ/CMC FHIR pilot are *MedicationKnowledge*, *Substance*, and *PlanDefinition*, which are previously existing resources integrated within FHIR that were not developed specifically for the pilot.[35] *MedicationKnowledge* is a Resource block that collates descriptive information about drug product (names, dosage forms, dosage strengths) and *Substance* is, accordingly, a Resource used to manage nonquantitative drug substance information (names, company codes). *PlanDefinition* is an adaptable Resource block that can be used to describe actionable events within a plan—meaning that it provides definitions for when specific actions are performed.[36] In this case, *PlanDefinition* is being used to group and define quality specifications, which are effectively planned actions (analytical tests) that must fall within fixed definitions or ranges. *PlanDefinition* can link out to either *MedicationKnowledge* or *Substance* to define specifications for either drug product or substance, respectively. Within *PlanDefinition*, the primary action is referred to as Test, which can be used to refer to categories of testing methods for determining substance or product identity, potency, physical state descriptions and properties, biological and chemical properties, and impurities. The action within the resource can have its own identity descriptors, references to other elements, titles, extensions, and "reasons," which provide context for why the action is being performed, such as for stability testing and batch release. Actions can also have subactions within (represented as *action action*). Within the PQ/CMC pilot, these subactions are known as "Stages," which provides structure for inputting multiple timepoints or replicates for Acceptance Criteria data. Within the Resource framework, Acceptance Criteria are defined as the *action action goal*—or the ideal specifications criteria set for the analytical testing methods for each "Stage." Once the logic and relationships between Stages are established, *Test* output data can be defined in *values*. *Interpretation codes* can then describe the relationship between observed *values* and the acceptance criteria. *Interpretation codes* are linked to *literal text*, which allows for text-based narrative description of the acceptance criteria defined in the *interpretation code*. Figure 3 provides a simplified graphical representation of quality specifications structured in FHIR format in the *PlanDefinition* Resource as detailed previously.

QOS Integration With KASA & PQ/CMC

The QOS appears in module 2 of the eCTD and as a result is a current requirement for NDA, BLA, and ANDA submissions in the United States and in other ICH member countries. The QOS provides a summary of quality data to provide context and assist the reviewer in understanding information presented in module 3. While the QOS has been a requirement since the introduction of the eCTD, in 2018 the FDA authored a white paper setting agency expectations for QOS content as there was insufficient guidance regarding its contents. The white paper indicates the agency's desire to further leverage the efficacy of the QOS as a tool.[16] The FDA indicated that the QOS should facilitate the reviewer's ability to clearly relate quality information to potential risks to patient safety, while also briefly summarizing risk mitigation and control plans.

The FDA has affirmed that the QOS does not alter or replace module 3, as it is meant to serve as an accompaniment.[30] Similarly, the KASA, PQ/CMC, and QOS initiatives can be complementary and coexist within the eCTD and beyond. For example, it can be envisioned that the QOS could provide summary data in a structured format that can refer directly to and provide context for raw PQ/CMC data entered in module 3. Some aspects of module 3, as it is currently defined, would still require traditional narratives but this kind of structured format for cross-referencing data has definite advantages in the context of data integrity and ability to quickly incorporate new data as it becomes available.

SCM Implementation Challenges

While KASA, PQ/CMC, and QOS have great potential for long-term benefit on behalf of industry, regulators, and patients, there are significant challenges to overcome in ensuring the effective implementation of SCM systems. Although KASA and PQ/CMC are being promoted as time-savers for all stakeholders involved, the necessary steps for establishing the infrastructure and systems needed for collecting and structuring data will be lengthy, be complex, and require substantial resources, particularly from industry. In addition, these initiatives partially destabilize significant

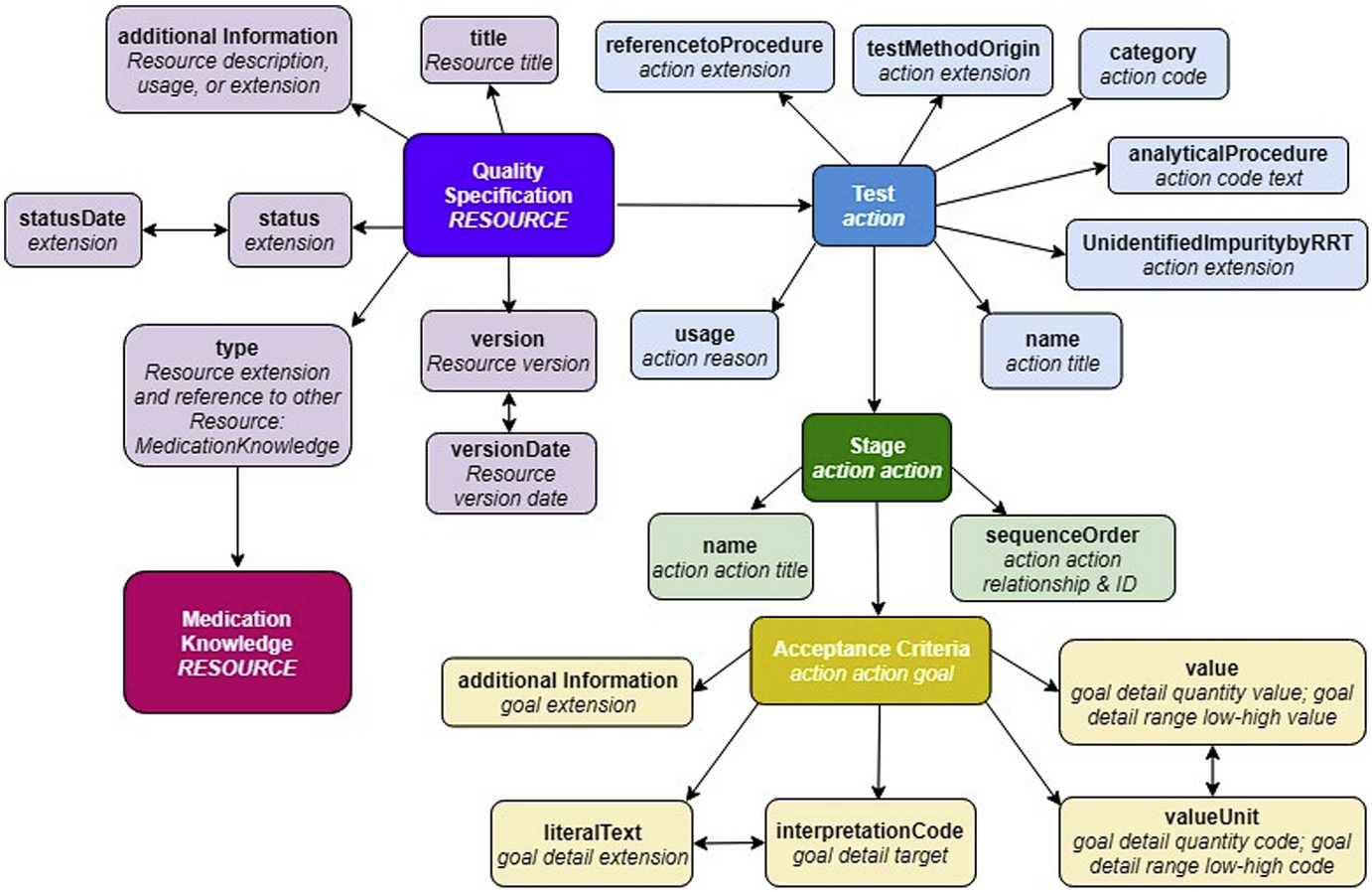

Figure 3. The FDA's PQ/CMC pilot uses HL7's FHIR data standards framework to structure aspects of CMC documents. The figure depicts a theoretical framework for the organization of specification data in the defined Resource element, PlanDefinition. It also describes how the specification Resource (PlanDefinition) can connect to other data elements, such as MedicationKnowledge, which contains drug product information. Created based upon FHIR PQ/CMC Implementation Draft documents.[34-36]

efforts toward global harmonization of regulatory submission content and format. While SCM solutions similar to those discussed in this article would transform and modernize industry's data collection methodology, advance opportunities for data mining, and promote the development of machine learning algorithms to improve manufacturing efficiency, the added workload in managing increasingly segregated global regulatory requirements could be substantial enough to negate any potential time saving, particularly for postcommercialization filings.

Although KASA and PQ/CMC may result in a more effective review system for the FDA in the United States, the efficiency of pharmaceutical companies to produce high-quality, globally available therapeutics and implement needed manufacturing changes may suffer significantly in the absence of harmonization, increasing filing time on a global scale, and eventually slowing patient access to medicines. Notably, the EMA has made progress in using structured approaches for managing regulatory submission data. Article 57 (2) of Regulation, issued in 2012, required the EMA to assemble a repository of all approved medicines for use in humans in Europe, which also subsequently required companies to provide structured electronic information on their products via a Web-based data portal to support the EMA's repository-building efforts. This portal system, known as EVWeb, was a crucial component in establishing an electronic exchange system between regulatory agencies and industry and may be the first step toward expanding requirements for structured electronic data submission.[37] Accordingly, the long-term success of the EMA's structured data initiatives, such as EVWEB, holds significant implications for the future success of the FDA's KASA. Ideally, these conceptually similar initiatives would establish concurrence for application to submissions within the EU and US, providing additional incentives for industry to invest time and resources into developing SCM systems.

Making Connections: Future Goals of SCM for CMC Data

Data Integrity

Beyond KASA and PQ/CMC, the integration of SCM solutions to manage holistic pharmaceutical data—from clinical to CMC, nonclinical, discovery, and beyond—could have profound positive impact upon business operations overall across the global pharmaceutical industry. Introducing SCM has great potential to aid in authoring high-quality, accurate, consistent documents and significantly improve data integrity throughout product lifecycles. Maintaining data integrity should be of particular concern to all industry stakeholders, as deficiencies in ensuring data integrity are cited as a frequent cause of violations, leading the FDA to issue warning letters, consent decrees, and import alerts.[38]

Employing SCM solutions for data management will allow for greater confidence in maintaining data integrity by eliminating many of the current review and verification cycles that are a vital aspect of the current regulatory document maintenance workflow. As discussed previously in greater detail, if a data point changes, it will only need to be edited once within its content block, and the

change will be automatically applied to all documents that the edited content block is linked to, across a central data repository. Accordingly, this also means that the data will only need to be verified within the content block, rather than made individually across all documents in which the change appears, which can be cumbersome and difficult to track.

Increasing Automation and Opportunities for Machine Learning

In addition to maintaining data integrity, integration of SCM solutions within data collection and authoring systems in the biotechnology and pharmaceutical industries is a productive initial step toward increasing overall automation, allowing for opportunities to leverage process-improving techniques in machine learning. Other functions within the pharmaceutical industry have utilized machine learning to their great advantage. A particular project of interest within the drug discovery space is the creation of the Machine Learning Ledger Orchestration for Drug Discovery (MELLODDY), a collaborative effort between 11 pharmaceutical companies that utilizes preclinical data and chemical libraries from multiple companies, while maintaining confidentiality of assay specifics and company interests, to build models that aim optimize compound and target development.[39] A similar machine learning technology that facilitates both connectivity and confidentiality could be used to optimize manufacturing procedures and minimize risk, particularly as companies become increasingly interested in novel drug delivery systems and formulations, for which all collective experience is valuable, as these new technologies have comparably less historical data.

Enhancing automation is central to reaching the goal of operating at Industry 4.0 standards. Major components of Industry 4.0, defined as the upcoming revolution in automation in manufacturing, include human-machine communication and networks of Cyber-Physical systems, in which machines use software algorithms to communicate with a network of other machines and respond to feedback from humans.[40] The machines working in this system can be capable of improving processes over time. Currently, many operations within the pharmaceutical industry fall short of Industry 4.0 standards, as equipment requires significant manual input and is not versatile enough to allow for machine learning capabilities. Although the current IT initiatives appear to propose a submission-based system that shows a manually submitted, static view of captured data, a futuristic perspective of such an initiative may allow for continual, updated submissions of real-time data to various health authorities to enhance transparency and availability of data.

In the distant future, one could even envision a central submission point or portal for global health authorities, who could "reach in" to the secured portal and "pull" the data that they require for review, while also having access to information requests and approval status for that application from other regulators. This is favorable for regulators, who will have ongoing and real-time access to quality data, as well as for industry, as staff will not have to organize and restructure large dossiers of data, repetitively, for multiple rounds of review and multiple jurisdictions. Risk-based assessments could be performed essentially in real time concurrently by individual health authorities around the world. This also presents another possibility to leverage the principles outlined in ICH Q12 through the implementation of PQ/CMC by using the structured format of data managed through FHIR to establish and maintain PLCM documents and to enable industry to make informed changes to process data in real time. In this high-tech paradigm, lifecycle management could be built-in and the system will maintain itself without requiring manual formatting and reformatting into different documents.

AI Data Mining of Structured Data

Currently, AI technology and machine learning, in particular, works most efficiently with structured data, which is typically repetitive, pattern-based, and more amenable to the development of "training data sets" that are used to train computers to work with specific types of data.[41] However, AI is becoming increasingly adept at converting unstructured data, such as photographs, audio files, and more, into structured data, providing potentially dynamic advantages for adoption of SCM systems. For example, in 2017, Box, a cloud storage and content management service, began developing an image-recognition platform in collaboration with Google that will automatically add text "tags" to photos based on their content and enable transcription of text in images, allowing for machine-aided image identification and sorting of archived images.[42] In the context of CMC data, an image tagging system could work alongside image sensor technology, which could, for example, identify product defects (discoloration, solution turbidity, damaged vials, etc.). The images collected by the image sensor could then be cataloged, stored, and queried using SCM applications. Although these technologies, as described, are not presently in use and likely face numerous logistical barriers before successful implementation, creative opportunities to integrate a variety of AI strategies are continually increasing in number and should be taken advantage of to enhance the quality and efficiency of pharmaceutical manufacturing and analytical procedures, in addition to simplifying regulatory oversight responsibilities.

Facilitating Direct Connections to Health Care

The health care industry generates exabytes of data, which can be used meaningfully to track product quality, collect real-world data, design biomarkers, and more. For example, the regional U.S. health network system, Kaiser Permanente, has generated between 25 and 44 petabytes of data from its 9 million members.[43] However, the pharmaceutical industry currently has no way of meaningfully harnessing these data and is relatively disconnected from the health care system outside of its own clinical and pharmacovigilance studies. Therefore, the use of FHIR in FDA's PQ/CMC program may confer a significant advantage in linking the quality aspects of the pharmaceutical industry with patient care, as patient data are already being tracked using this system. Linkages between the pharmaceutical industry and patient care have the potential to enhance adverse event tracking, which is currently submitted into separate databases, allowing for better understanding and resolution of potential quality-related issues in real time. Harnessing linkages to telehealth and digital health tools can also help to inform regulatory and quality decision making by providing real-time data outside of clinical spaces, such as physician's offices and hospitals.

Conclusions

SCM solutions offer dynamic benefits for managing regulatory submissions, while also leveraging potential for enhancing modernization by facilitating machine learning algorithms, data mining, and connections to real-time patient data for monitoring of safety and efficacy. Several pharmaceutical companies and industry working groups have made efforts to standardize and automate data collection. Such efforts have been met with some success in enhancing time saving, ensuring data reusability, and integrating with existing authoring systems for ideal usability and user friendliness. However, most of these prior ventures toward SCM systems have occurred in the context of clinical data management and labeling, with few applications for CMC-specific data. The FDA

is attempting to catalyze industry investment and participation in the collection and submission of structured CMC data through its KASA and PQ/CMC initiatives.

Although KASA and PQ/CMC conceptually represent a preliminary effort toward use of SCM for CMC data in the United States, many unique roadblocks and hurdles can slow or preclude the timely adoption of SCM on an international scale. Global harmonization of regulations plays a pivotal role in predicting the success of SCM implementation across industries and health agencies. If other regions do not agree to adopt similar standards and approaches as the FDA, companies may face significant burden in tailoring their applications to meet regional needs. Though the intended goal is to grant consumers earlier access to high-quality, safe, and effective pharmaceutical products, KASA and PQ/CMC may create a wider discrepancy between global regulatory applications, conceivably delaying global access to life-saving therapeutics. Currently, most ICH regions are leveraging the eCTD format with a small number of regions still working with paper submissions. Health authorities would need to revise their current operational plans and develop a complex IT framework to accommodate an SCM-submission format. Even regions that have formally adopted eCTD format will face initial high investment costs in developing the IT systems necessary to support these initiatives.

Ideally, SCM solutions can integrate as drivers for upholding ICH quality guidelines to promote global adaptation and foster industry-wide acceptance of an upgraded data management structure. There are similar concepts in KASA and ICH Q12 for organizing CQAs and ECs into predefined categories, suggesting a collective interest in increasing standardization of data and risk assessment. If successfully implemented, SCM may confer benefits to regulatory agencies and industry on a global scale by eliminating unnecessary paperwork, decreasing manual document interfaces, and allowing for easy comparison between data sets, thereby improving regulatory decision making and knowledge across product lines, which translates to providing a direct benefit to patients in need of therapies worldwide.

Acknowledgments

The authors would like to thank Simon Hotchin, Cenk Undey, and Linda Narhi for their helpful suggestions.

References

1. Friedman R. Manufacturing, measurements & quality systems: seeking order in complexity. 2009. Available at: http://pqri.org/wp-content/uploads/2015/08/pdf/Presentation_Friedman.pdf. Accessed June 3, 2019.
2. Ekins S. The next era: deep learning in pharmaceutical research. *Pharm Res*. 2016;33(11):2594-2603.
3. China National Medical Products Administration (NMPA). Reply to Recommendation No. 7413 of the Fifth Session of the Twelfth National People's Congress (Proposal on Further Deepening the Reform of the Drug Review and Approval System) [对十二届全国人大五次会议第7413号建议的答复（关于进一步深化药品审评审批制度改革的建议）]. 2017. Available at: http://www.nmpa.gov.cn/WS04/CL2140/318076.html. Accessed June 19, 2019.
4. International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. M4: the common technical document. Available at: https://www.ich.org/page/ctd. Accessed January 23, 2020.
5. Yang W, Wan L. China gets ready to launch eCTD, as companies face a challenging learning curve. 2019. Available at: https://blogs.dxc.technology/2019/05/16/china-gets-ready-to-launch-ectd-as-companies-face-a-challenging-learning-curve/. Accessed June 19, 2019.
6. Allotrope Foundation. Available at: https://www.allotrope.org. Accessed June 27, 2019.
7. QUDT. Available at: https://qudt.org. Accessed June 27, 2019.
8. Pistoia alliance. Available at: https://www.pistoiaalliance.org. Accessed June 27, 2019.
9. *Electronic Code of Federal Regulations. 21.314.70: Supplements and Other Changes to an Approved NDA*. US Food and Drug Administration. Government Publishing Office; 2016. Available at: https://www.federalregister.gov/documents/2004/04/08/04-7532/supplements-and-other-changes-to-an-approved-application. Accessed June 6, 2019.
10. Yamamoto G. Consideration point of post-approval change. 2016. Available at: https://www.pmda.go.jp/files/000215552.pdf. Accessed June 6, 2019.
11. DitaExchange. *How Structured Content Management (SCM) Is Revolutionizing the Life Sciences Industry*; 2015. Available at: https://ditaexchange.com/wp-content/uploads/2015/06/SCM-White-Paper.pdf. Accessed June 4, 2019.
12. US Food and Drug Administration. Fast track, breakthrough therapy, accelerated approval, priority review. Available at: https://www.fda.gov/patients/learn-about-drug-and-device-approvals/fast-track-breakthrough-therapy-accelerated-approval-priority-review. Accessed June 16, 2019.
13. Richardson E, Daniel G, Joy DR, et al. Regional approaches to expedited drug development and review: can regulatory harmonization improve outcomes? *Ther Innov Regul Sci*. 2018;52(6):793-798.
14. Cauchon NS, Oghamian S, Hassanpour S, Abernathy M. Innovation in chemistry, manufacturing, and controls-A regulatory perspective from industry. *J Pharm Sci*. 2019;108(7):2207-2237.
15. Bray T, Paoli J, Sperberg-McQueen MC, Maler E, Yergeau F. Extensible markup language (XML) 1.0. Available at: https://www.w3.org/TR/xml/. Accessed June 5, 2019.
16. Harrison N. The Darwin Information Typing Architecture (DITA): applications for globalization. In: *IPCC 2005 Proceedings: International Professional Communication Conference*. 2005:115-121. Available at: https://ieeexplore.ieee.org/document/1494167.
17. ArborSys. Implementing a structured content management solution-for product labeling in a pharmaceutical company. Available at: http://www.arborsys.com/PDFs/implementing-a-structured-content-management-solution-for-product-labeling-in-a-global-pharmaceutical-company.pdf. Accessed June 4, 2019.
18. Zepper N. Finding flexibility in a rigid world. In: *Intelligent Content on Life Sciences and Healthcare Conference*; 2014. Available at: https://vimeo.com/140185439. Accessed June 5, 2019.
19. Porter Booker TK, Cruz Naranjo M, Horowitz A. *A Case Study in Structuring Clinical Content and Structured Content Management (SCM)*. San Diego, CA: DIA 2019; 2019.
20. Affleck J, Bahl B, Ranganathan V. Sanofi – the ENCORE platform: enterprise content optimization & reuse engine. 2012. Available at: https://ditaexchange.com/wp-content/uploads/2015/04/DX_ShareFEST-2012-Affleck-Bahl-Ranganathan.pdf. Accessed June 5, 2019.
21. Ehret-Kayser. *Connecting Data to Power the Pharma Supply Chain*; 2019. Available at: https://www.marklogic.com/blog/connecting-data-to-power-the-pharma-supply-chain/. Accessed July 8, 2019.
22. Gottlieb S. Budget matters: infrastructure to support robust generic drug competition. 2018. Available at: https://www.fda.gov/news-events/fda-voices-perspectives-fda-leadership-and-experts/budget-matters-infrastructure-support-robust-generic-drug-competition. Accessed May 21, 2019.
23. Food and Drug Administration. *Annual Report: One Quality Voice. 2017*; 2017. Available at: https://www.fda.gov/media/111436/download. Accessed May 30, 2019.
24. Rosencrance S, Raw A, Smith D, Slack MA. *FDA's New Initiative: KASA*; 2019. Available at: https://pqri.org/wp-content/uploads/2019/04/PQRI_KASA-Presentation_V4.pdf. Accessed May 30, 2019.
25. Brennan Z. *End of the eCTD? FDA Pushes for New KASA System to Improve Assessments. RAPS Regulatory Focus*; 2018. Available at: https://www.raps.org/news-and-articles/news-articles/2018/9/end-of-the-ectd-fda-pushes-for-new-kasa-system-to. Accessed June 21, 2019.
26. US Food and Drug Administration. The generic drug review Dashboard: as of July 1, 2016. 2016. Available at: https://www.fda.gov/media/99155/download. Accessed June 21, 2019.
27. US Food and Drug Administration. *Activities Reports of the Generic Drugs Program - FDARA Title VIII Sections 807 and 805 (FY 2018)*; 2018. Available at: https://www.fda.gov/drugs/generic-drugs/activities-reports-generic-drugs-program-fdara-title-viii-sections-807-and-805-fy-2018. Accessed June 21, 2019.
28. Yu LX, Raw A, Wu L, Capacci-Daniel C, Zhang Y, Rosencrance S. FDA's new pharmaceutical quality initiative: knowledge-aided assessment & structured applications. *Int J Pharm*. 2019;1:100010.
29. Schmuff NR. 2nd public meeting on PDUFA VI electronic submissions and data standards: structured PQ/CMC. 2019. Available at: https://www.fda.gov/media/124614/download. Accessed June 21, 2019.
30. US Food and Drug Administration. *A Regulatory Perspective on the Quality Overall Summary: Putting the Pieces Together*; 2018. Available at: https://www.fda.gov/media/110657/download. Accessed June 7, 2019.
31. International conference on harmonisation of technical requirements for registration of pharmaceuticals for human use. ICH harmonised guideline: technical and regulatory considerations for pharmaceutical product lifecycle management Q12. Available at: https://www.ich.org/page/quality-guidelines#12. Accessed January 23, 2020.
32. US Food and Drug Administration. *Pharmaceutical Quality/Chemistry Manufacturing and Controls (PQ/CMC) Data Elements and Terminologies*; 2017. Available at: https://www.regulations.gov/document?D=FDA-2017-N-2166-0002. Accessed May 21, 2019.
33. Health Level 7. FHIR Overview. 2018. Available at: https://www.hl7.org/fhir/overview.html. Accessed June 21, 2019.
34. Health Level 7. Project summary for pharmaceutical quality/chemistry, manufacturing and controls (PQ/CMC) (PSS in confluence). 2019. Available at:

http://www.hl7.org/Special/committees/rcrim/projects.cfm?action=edit&ProjectNumber=1537. Accessed June 21, 2019.
35. Health Level 7. FHIR implementation guide home page: PQCMC proof of concept. 2019. Available at: https://github.com/HL7/FHIR-pq-cmc. Accessed January 23, 2020.
36. Health Level 7. Resource PlanDefinition – content. 2018. Available at: https://www.hl7.org/fhir/plandefinition.html. Accessed June 26, 2019.
37. Grønning N. Data management in a regulatory context. *Front Med (Lausanne)*. 2017;4:114.
38. US Food and Drug Administration. *Data Integrity and Compliance with Drug cGMP: Questions and Answers Guidance for Industry*; 2018. Available at: https://www.fda.gov/media/119267/download. Accessed June 20, 2019.
39. Burki T. Pharma blockchains AI for drug development. *Lancet*. 2019;393(10189):2382.
40. Brettel M, Friederichsen N, Keller M, Rosenberg M. How virtualization, decentralization and network building change the manufacturing landscape: an Industry 4.0 Perspective. *Int J Mech Ind Sci Eng*. 2014;8(1):37-44.
41. Pinto Dos Santos D, Baeßler B. Big data, artificial intelligence, and structured reporting. *Eur Radiol Exp*. 2018;2(1):42.
42. Frank BH. Box taps Google's AI for image recognition. 2017. Available at: https://venturebeat.com/2017/08/17/box-taps-googles-ai-for-image-recognition/. Accessed July 8, 2019.
43. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health Inf Sci Syst*. 2014;2:3.