Challenges of Trustable AI and Added-Value on Health
B. Séroussi et al. (Eds.)

doi:10.3233/SHTI220578

Towards a Computational Approach for the Assessment of Compliance of ALCOA+ Principles in Pharma Industry

Marta DURÁ[a,] and Ángel SÁNCHEZ-GARCÍA[a], Carlos SÁEZ[a], Fátima LEAL[b,c], Adriana E. CHIS[c], Horacio GONZÁLEZ-VÉLEZ[c], and Juan M. GARCÍA-GÓMEZ[a]
[a] *Biomedical Data Science Lab, Instituto Universitario de Tecnologías de la Información y Comunicaciones, Universitat Politècnica de València, Camino de Ver s/n, Valencia 46022, Spain*
[b] *REMIT, Universidade Portucalense, R. Dr. António Bernardino de Almeida 541, 4200-072 Porto, Portugal*
[c] *Cloud Competency Centre, School of Computing, National College of Ireland, Mayor Street, IFSC, Dublin D01 K6W2, Ireland*

Abstract. The pharmaceutical industry is a data-intensive environment and a heavily-regulated sector, where exhaustive audits and inspections are performed to ensure the safety of drugs. In this context, processing and evaluating the data generated in the manufacturing lines is a relevant challenge since it requires compliance with pharma regulations. This work combines data integrity metrics and blockchain technology to evaluate the compliance-degree of ALCOA+ principles among different levels of drug manufacturing data. We propose the DIALCOA tool, a software to assess the compliance-degree for each ALCOA+ principle, based on the assessment of data from manufacturing batch reports and its different levels of information.

Keywords. Data integrity, ALCOA+ compliance, pharma manufacturing industry

1. Introduction

The pharma manufacturing industry is a data-intensive environment that generates large amounts of distributed data regularly accessed by different internal and external stakeholders including international and national regulatory bodies. However, this is hardly a new problem: since the early 1960s when the initial Good Manufacturing Practices (GMPs) [1] for finished pharmaceuticals were published, distinct regulatory bodies have assembled a considerable number of guidelines pertaining to data integrity in pharma manufacturing.

Despite the pharmaceutical industry has consistently improved its manufacturing processes in compliance with good manufacturing practices, it is well documented that falsification of medicines continues[2] and has led to disastrous consequences worldwide[3]. Consequently, different organizations have proposed standards, measures, and protocols to avoid these falsifications. The EU Falsified Medicines Directive[4] introduces harmonized European measures to fight these medicine falsifications and ensure that medicines are safe and that the trade in medicines is rigorously controlled."

Such obligatory safety features, legal framework, and record-keeping requirements have arguably imposed stricter controls for the manufacturing of medicines.

In this context, the gold standard adopted by the pharmaceutical industry is "Data Integrity and Compliance with current Good Manufacturing Practices", defined by the FDA[5], which defines the term "ALCOA+" as a set of principles that should be followed throughout the data life cycle for achieving data integrity. These principles stand that data should be Attributable, Legible, Contemporaneous, Original, and Accurate. Moreover, good documentation practices require that the records are Complete, Consistent, Enduring, and Available.

This work proposes a computational approach for the assessment of these nine ALCOA+ principles among the data generated during the process of drug manufacturing, in order to provide a quantitative measurement of data integrity compliance level. This work has been developed under the Smart Pharmaceutical Manufacturing project (SPuMoNI), a European research project launched by CHIST-ERA pathfinder programme. SPuMoNI consortium includes industry partners, such as a Contract Manufacturing Organisation (CMO), which has been a real scenario for the development of this work.

2. Methods

2.1 Pharma manufacturing reports

At CMOs, the process of manufacturing a particular drug is performed by following the Recipe, which is the protocol that describes in detail the fabrication process of the drug. It is composed of a set of Phases, and each Phase is composed of a set of Instructions. An Instruction is a single action implemented within the manufacturing process. There are various types of Instructions, such as setting a mixing machine or verifying the quantities of raw materials. All this information must be extensively documented and is expected to be ALCOA+ compliant, since regulatory bodies or any other auditor could require an exhaustive revision at any time.

We propose to structure this manufacturing data in what we call a batch Report[6]. At the main level of a Report, the attributes are related to the batch information, such as the batch code, the Recipe code, and the Qualified Person; which is responsible for assuring the quality of the manufactured drugs that will be available on the market. Furthermore, a Report contains the list of materials used for the process as well as the row data produced in the production line. This information is organised in the Report following the Recipe structure: a) a list of Phases that contains a set of Instructions; b) each Instruction item includes a list of parameters to be controlled during the execution and the data recorded during the process (*e.g.*, temperatures or mixing speeds).

2.2 ALCOA+ principles assessment

Following the definitions of ALCOA+ and data integrity methods[6], we have defined a set of metrics (Table 1) for evaluating the compliance of each principle in a batch Report. These metrics are implemented in the tool proposed in this work (Results Section) that we have named "DIALCOA" (**D**ata **I**ntegrity **ALCOA** assessment).

Table 1. ALCOA+ Principles definition and their proposed metrics for assessing its compliance in batch manufacturing Reports

	ALCOA+ Principle	Proposed Metric
Attributable	All data must be attributable to the person generating the data including who performed an action and when.	Assessed by measuring the amount of data which have been assigned to the person who did the collection and the identification of the person responsible of the report Attributable score is the percentage of data which has been recorded by the staff who has collected it.
Legible	All data must be legible and permanent. Ensuring records are legible	Assessed among all Report fields by the quantification of measurements that comply with these three specifications: - Data must be electronic and use UTF-8 format - Decimal numbers must use the same format - Free texts must use words present in language dictionaries Legible score is the percentage of data that is compliant with the three specifications.
Contemporaneous	Record the data at the time it is performed. Date and time stamps should flow in Date and time stamps should flow considering the execution order to prove the data credibility.	Assessed by verifying and counting the Report fields which include the date and time of data creation. Contemporaneous score is the percentage of data that includes its timestamp.
Original	The preservation of original records, to verify the authenticity of the data.	Assessed by verifying that the Report has not been adulterated. To do so, DIALCOA relies on a blockchain Smart Contract, where the original version of the Report is stored. Original score is 100 if data is original or 0 if any field of the Report has been adulterated.
Accurate	Data should be free from errors. Editing should only be performed by using the principles of GDPs.	Assessed by range checks and outlier detection methods. Those numerical fields of the Report are checked with the expected range of acceptable values in the Recipe, and to detect outlying behaviour. Accurate score is the percentage of numerical fields of the Report that satisfy the rules above.
Complete	All data must be retained from the creation of the documentation. Deletion or removal of data must not take place.	Assessed by checking that all expected fields in the Report are fulfilled. Complete score is the percentage of fulfilled fields among the Report.
Consistent	Data must be coherent with the expected information in time.	Consistency principle can be assessed by evaluating time consistency, counting the tracking start date that is earlier than tracking end date for all the Report. Consistency score is the percentage of data that is compliant with this time consistency rule.
Enduring	Store systems are running during all life of data.	Assessed by requesting a certified expiration date of the Report. Enduring score will be 100 if the expiration date is included and updated and 0 if it is missing or if it has expired.

Available	Data must be accessible at any moment it is request.	Assessed by requesting a certified expiration date for accessing to the Report. Available score will be 100 if the expiration date is included and updated and 0 if it is missing or if it has expired.

2.3 Blockchain private network

The DIALCOA tool is connected to a blockchain private network that is composed of a private Ethereum network infrastructure[7]. When a Report is uploaded to DIALCOA, an originality assessment is performed by uploading a new batch record to the Ethereum network as a smart contract and verifying the uniqueness of all its data. Based on the previously uploaded reports on the Ethereum network, the originality score is calculated evaluating the uniqueness of the new data by comparing it with the existing stored information.

3. Results

This work presents the first steps towards a computational approach to assess the compliance of ALCOA+ principles within batch manufacturing data. This software implements the proposed metrics described in Table 1 to be used by the Qualified Person at the CMO to monitor the integrity of the data that have been generated.

This software can be installed in the pharma manufacturing plant systems to access production data and batch Reports. Additionally, a private blockchain network should be installed in order to ensure the traceability of the ALCOA+ assessments and the Original principle evaluations. Figure 1 outlines the information workflow and the connection among elements.

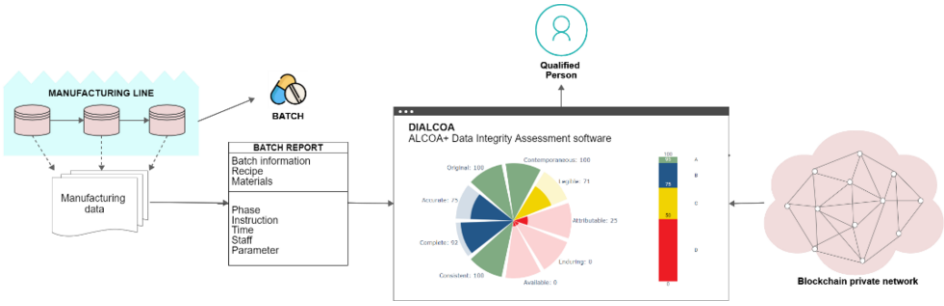

Figure 1. Workflow of DIALCOA tool in pharma manufacturing process

DIALCOA shows a global view of the nine ALCOA+ principles scores in a pie chart, including a color scale for the limits of compliance (Figure 1). Moreover, the user can explore the detailed analysis of each principle assessment. This is possible since the software is able to detect and plot the potential data integrity conflicts which are causing scores lower than 100. Hence, the user can easily identify which Report data present data integrity issues for each ALCOA+ principle.

Discussion

The proposed system feasibly supports the compliance of ALCOA+ principles by evaluating batch Reports through data integrity metrics. To achieve a higher readiness level, an evaluation of the proposed tool in the pharma shop-floor environment is being performed. As future work, we aim to validate DIALCOA tool in a real pharma manufacturing environment.

4. Conclusion

The pharmaceutical industry is a data-intensive and heavily regulated domain. Its manufacturing lines continuously generate large amounts of data that must be collected and have to be ALCOA+ compliant. This industry requires effective solutions to improve its manufacturing process in terms of data integrity compliance. Towards this scenario, we propose a novel tool for assessing the compliance of ALCOA+ principles within batch manufacturing reports.

Acknowledges

We thank the consortium of Smart Pharmaceutical Manufacturing project (https://www.spumoni.eu/) for enabling the development of this work in a real scenario of pharma industry. We also thank the co-funding support of European Union and Spanish Agencia Estatal de Investigación (PCI2019-103783) for the development of SPuMoNI project.

References

[1] US Federal Register–Part 133, 28 FR 6385, 20/Jun/63. Drugs, Current Good Manufacturing Practice in Manufacture, Processing, Packing or Holding. American Journal of Hospital Pharmacy. 1964 09;21(9):398–401

[2] D. McManus, B. D. Naughton, A systematic review of substandard, falsified, unlicensed and unregistered medicine sampling studies: a focus on context, prevalence, and quality, BMJ Global Health 5 (8) (2020) e002393. doi:10.1136/bmjgh-2020-002393.

[3] M. S. Rahman, N. Yoshida, H. Tsuboi, N. Tomizu, et al., The health consequences of falsified medicines-a study of the published literature, Tropical Medicine & International Health 23 (12) (2018) 1294–1303. doi:10.1111/tmi.13161.485

[4] European Parliament–Council of the European Union, Directive 2011/62/EU of the European Parliament and of the Council of 8 June 2011 amending Directive 2001/83/EC on the Community code relating to medicinal products for human use, as regards the prevention of the entry into the legal supply chain of falsified medicinal products, Official Journal of the European Union 54 (2011) Document 32011L0062. doi:10.3000/17252555.L_2011.174.eng

[5] US Department of Health and Human Services. Data Integrity and Compliance With CGMP Guidance for Industry. Silver Spring: Food and Drug Administration; 2016. Pharmaceutical Quality/Manufacturing Standards (CGMP)

[6] Leal F, Chis AE, Caton S, González-Vélez H, García-Gómez JM, Durá M, et al. Smart Pharmaceutical Manufacturing: Ensuring End-to-End Traceability and Data Integrity in Medicine Production. Big Data Research. 2021;24:100172. (Open Access).

[7] Leal F, Chis AE, González-Vélez H. Performance Evaluation of Private Ethereum Networks. SN Computer Science. 2020;1(5):285