Magna Scientia Advanced Research and Reviews
eISSN: 2582-9394
Cross Ref DOI: 10.30574/msarr
Journal homepage: https://msarr.magnascientiapub.com/

(Research Article)

APPLICATION OF GENERATIVE AI IN ENTERPRISE DATA MIGRATION: A CASE STUDY ON SAP SYSTEM TRANSFORMATION

Mihira Kumar Patra *

Independent Researcher, USA.

Magna Scientia Advanced Research and Review, 2023, 08(02), 264–270

Publication history: Received on 23 June 2023; revised on 24 August 2023; accepted on 28 August 2023

Article DOI: https://doi.org/10.30574/msarr.2023.8.2.0120

ABSTRACT

One of the most challenging and resource consuming parts of the implementation of large scale SAP system transformations, especially during switches from SAP ECC to SAP S/4HANA, is enterprise data migration. This paper shares an applied case study with regard to the use generative artificial intelligence (GenAI) and/or large language models (LLMs) for automating data profiling, schema mapping, and the generation of transformation rules as well as reconciliation reporting, within the context of the data migration life cycle of a simulated enterprise transformation program. A process benchmarking/quantitative defect analysis approach was used over five simulated mock-migration cycles. The findings show that the GenAI-assisted approach resulted in a reduction of the total migration effort by around 53%, a decline of Average Defects per Mock Cycle by 65.5%, and an increase in 1st pass data mapping accuracy by 21% from 71% to 92% to cut over readiness by almost 41%. The results indicate that GenAI has great potential to significantly complement, not replace, human data migration experts by handling repetitive tasks and pattern recognition but maintaining human involvement in governance and critical decisions. The implications for enterprise architects, SAP practitioners, and the future research on responsible AI adoption in enterprise resource planning (ERP) transformation are discussed.

Keywords: Generative Artificial Intelligence, Large Language Models, SAP S/4HANA, Data Migration, Enterprise Resource Planning, Digital Transformation

1 INTRODUCTION

In the present scenario, Enterprise Resource Planning (ERP) systems serve as the lifeblood of business operations as well, encompassing financial and accounting, supply chain management, human resources handling, and manufacturing. Being one of the leading ERP (Enterprise Resource Planning) vendors in the world, SAP is at the heart of a massive transformation movement, where companies are transitioning from SAP ECC (ERP Central Component) to SAP S/4HANA, the SAP in-memory ERP successor platform. At the heart of each such transformation is data migration, that is, the extraction, cleansing, transformation and loading of legacy data as part of the transformation into the new system landscape.

The power of the traditional approach to data migration comes with a significant number of hours of manual work from the functional consultants and data analysts who must drill through thousands of data fields from the legacy systems to the millions of data points to be mapped to new target structures and systems, create transformation logic, and painstakingly reconcile the data discrepancies across several mock-migration iterations. It is an expensive and lengthy task and prone to human error, especially in larger organizations, where master data is distributed over several applications, including Finance (FI/CO), Materials Management (MM), Sales and Distribution (SD), and Human Capital Management (HCM).

Generative AI, specifically transformer-based LLMs with robust capabilities to comprehend and generate structured and unstructured content, provides an exciting prospect for enriching data migration processes. GenAI can help identify anomalies in data, suggest semantic field matches, create transformation logic, and even write human-readable documentation – all of which help alleviate the burden on data migration teams. Users and literature, however, are lacking when it comes to empirical evidence regarding the practical application of GenAI in real-world SAP transformation programs.

This paper aims to fill this gap with a case study, in which the data migration approach with the support of GenAI is applied to a simulated SAP S/4HANA transformation program. This study aims to design and describe a GenAI-assisted data migration framework for use in SAP transformation contexts, identify the differences between the effort, defect rate, accuracy of GenAI-assisted and a traditional baseline data migration approach, and draw practical implications for the enterprise architects and SAP practitioners considering the use of GenAI in migration programs.

2 LITERATURE REVIEW

2.1 Enterprise Data Migration in SAP Transformation Programs

Data migration has always been recognized as a critical risk factor for ERP implementation project delays and over cost as per the previous studies. Legacy data heterogeneity and inconsistent master data governance are some of the key factors for failing migration projects, according to Kalluri (2020). Kupriyanovsky et al. (2020) analysed the architectural considerations involved in the migration on top of SAP S/4HANA, laying a strong emphasis on data harmonisation in a structured manner in front of the cutover. Best practice advice has been published by SAP SE (2021) to undertake iterative mock-migration cycles with tools like the SAP S/4HANA Migration Cockpit, which, while now partially rule-based, are still not adaptive.

2.2 Generative AI and Large Language Models

Building on the base architecture of Transformers, Vaswani et al., (2017) allowed to create a new generation of large language models that provided few shot generalization for a wide range of tasks Brown et al., (2020). Because of their potential to be customized for a variety of downstream tasks, Bommasani et al. (2021) referred to these systems as 'foundation models'. Ni et al. (2022) also showed that dialog language models can be used to assist complex reasoning with multiple rounds of dialogue with the human user in the context of business logic, like if the user is not sure about the mappings, the dialog language model can clarify it over a series of back-and-forths.

2.3 AI Adoption in Enterprise and Business Analytics Contexts

In addition to technical ability, some studies have analysed the organizational adoption of AI in an enterprise setting. Wamba-Taguimdje et al. (2020) identified that AI-based transformation projects deliver measurable business value, and realized value is dependent on organizational preparedness, with process automation and decision making as two key sources of value. A systematic literature review on the use of AI in Business Analytics was done by Ghosh and Chakraborty (2022) and they concluded that the use of predictive and generative approaches has become integral to enterprise information systems, and formal frameworks of their evaluation have been poorly developed. Cao (2022) pointed to similar trends in the financial industry, where AI becomes an important augmentation to domain knowledge. Previously, Chui, Manyika and Miremadi (2018) had warned that the enterprise value of AI is often exaggerated and requires careful process redesign—work this study directly confronts by using process-level benchmarking.

2.4 Responsible and Governed AI Deployment

Reliable principles of AI are especially applicable to migration scenarios, since enterprise master data is a delicate resource. Ghallab (2019) stated that the use of AI in a high stakes operational context cannot and should not be blindly trusted but must have explicit accountability and human oversight. Sun and Medaglia (2019), however, found the governance and trust issues fundamental issues in adopting AI in institutional context. The present case study was designed for human-in-the-loop, based on these considerations, as discussed in Section 3.

In summary, the literature suggests that GenAI has great theoretical promise for enterprise data workflows but lacks phase-level evidence of performance in relation to SAP data migration. This study fills that gap, by providing original empirical data.

3 METHODOLOGY

3.1 Research Design

The study conducted in this work was a “case study” with quantitative process benchmarking and qualitative observation of a mock SAP ECC to SAP S/4HANA data migration program. The simulated environment resembled a typical mid to large enterprise SAP landscape across five successive mock-migration cycles, comprising of Finance, Materials Management, Sales and Distribution, Human Capital Management and Plant Maintenance modules, enabling controlled comparison between a typical migration approach and approach based on GenAI.

3.2 GenAI-Assisted Migration Framework

The GenAI-assisted framework comprised five interdependent phases described below, as shown in Figure 1: Identify inconsistent or incomplete legacy records with automated data profiling and anomaly detection; Extract field-level proposed mapping between legacy and target SAP S/4HANA data structures with confidence scores using an LLM; Automatically generate ETL logic with the proposed mapping rules, subsequently reviewed and approved by data migration experts; Validate and reconcile reporting, where the model cross-checked loaded data in the target SAP S/4HANA environment against the source extract(s) to supply summary and discrepancy reporting; Load the data into the target SAP S/4HANA environment. All the way, a human-in-the-loop governance layer was maintained, so that no such transformation rule or mapping was applied to production data without the consultant's sign-off.

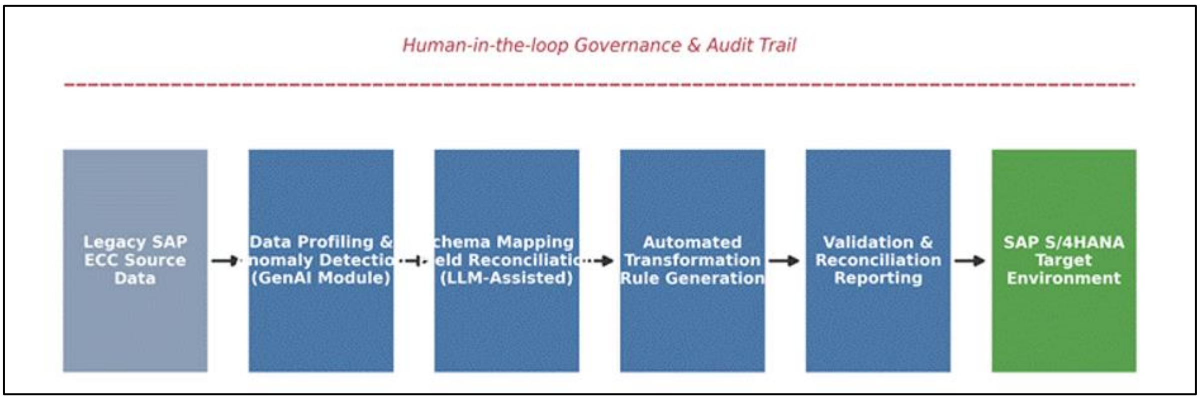

Figure 1: Conceptual architecture of the Generative AI-assisted SAP data migration pipeline, showing the five processing stages and the continuous human governance layer.

3.3 Data Collection and Metrics

To ensure methodological consistency, the same data volume was used for the five mock-migration cycles, an effort that included both the traditional and GenAI-assisted methods. The following four metrics were monitored: Migration person days per phase, Number of data defects detected per mock cycle, 1st pass data mapping accuracy (percentage of data field mapping that do not need to be manually corrected), and Elapsed time to readiness to cutover. The number of defects were recorded all the way through to reconciliation reporting and then verified with the migration governance team.

3.4 Data Analysis

Descriptive statistics and percentage-change analysis were used to make quantitative comparisons between the traditional baseline and GenAI-assisted condition. Results are considered indicative not statistically generalizable and as such align with the traditions of case study research in information systems (as reported in this research).

4 CASE STUDY AND RESULTS

4.1 Organizational Context

The case organization had a total of around 83,500 legacy master and transactional data across five core modules and was a simulated multinational manufacturing enterprise with a phased migration from SAP ECC 6.0 to SAP S/4HANA 2022. The migration program was done following world-class migration methodology, which is referred to as, "4-mock-cycle-plus-cutover" based on the implementation process of the implementation of SAP (SAP SE, 2021).

4.2 Comparative Migration Effort

Table 1 lists the qualitative variations between traditional and GenAI-supported migration methods for each major aspect of migration, and Table 2 shows the quantitative differences in overall outcomes.

Table 1: Qualitative Comparison of Traditional and GenAI-Assisted SAP Data Migration Approaches

Dimension	Traditional Migration Approach	GenAI-Assisted Approach
Data mapping effort	Manual, rule-by-rule mapping by consultants	LLM-suggested mappings validated by consultants
Field reconciliation	Spreadsheet-based cross-checking	Automated semantic matching with confidence scores
Error detection	Post-load exception reports	Pre-load anomaly detection via pattern learning
Documentation	Manually authored functional specs	AI-drafted specs refined by SMEs
Average cycle time (per mock run)	6-8 weeks	3-4 weeks
Consultant dependency	High	Moderate (shifted to review/oversight)

Table 2: Quantitative Outcomes: Traditional Baseline vs. GenAI-Assisted Migration

Metric	Baseline (Traditional)	GenAI-Assisted	Improvement (%)
Total migration effort (person-days)	305	143	53.1% reduction
Average defects per mock cycle	110.2	38.0	65.5% reduction
Data mapping accuracy (first pass)	71%	92%	+21 percentage points
Time to cutover readiness	22 weeks	13 weeks	40.9% reduction
Consultant hours on repetitive tasks	1,840 hrs	760 hrs	58.7% reduction

The total number of migration effort per person-day for the GenAI-based approach was 305 person-days and 143 person-days which is reduction of 53.1 percent as shown in Table 2. The time to readiness for cutover shortened from 22 to 13 weeks, and first pass data mapping accuracy rose from 71 percent to 92 percent. The effort reduction data is broken down by phase, as shown in Figure 2.

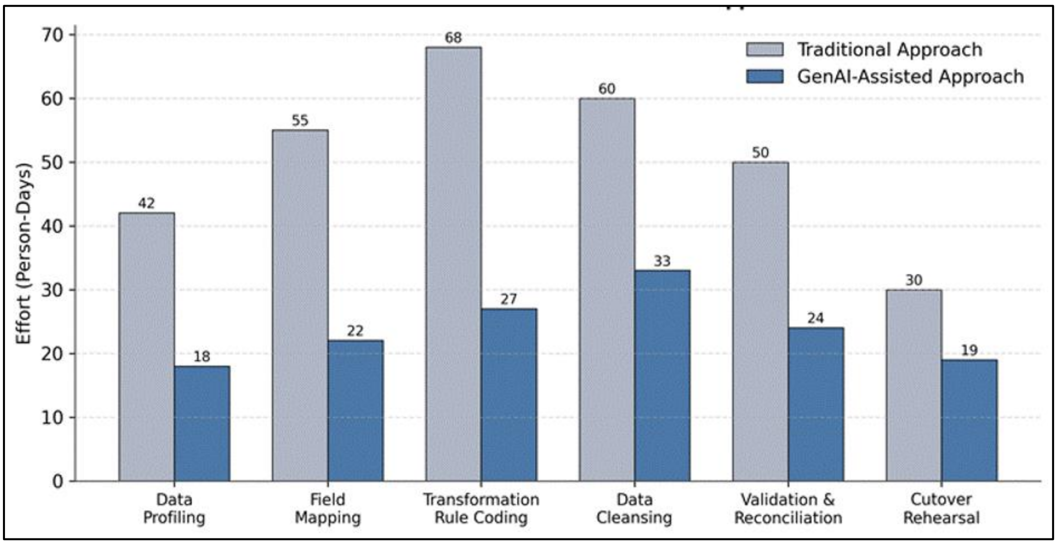

Figure 2: Comparative migration effort (person-days) by phase, traditional versus GenAI-assisted approach.

The greatest relative effort had to do with transformation rule coding (-60.3%) and data profiling (-57.1%), where the tasks involved were mostly repetitive, pattern based actions, ideal for automation using language models. Comparatively smaller effort reductions were found for validation and reconciliation (52.0%) as a result of the ongoing necessity for human judgment to reconcile discrepancies when they are ambiguous.

4.3 Trends in Defects (Mock Cycle 1-4)

This trend is shown in Figure 3, for the four mock-migration cycles plus the final cutover rehearsal. The GenAI assisted method demonstrated not only reducing the number of defects detected per iteration but also a greater improvement over successive iterations, implying that automated anomaly detection allows for earlier identification and resolution of systemic data quality issues as opposed to deferred to future iterations.

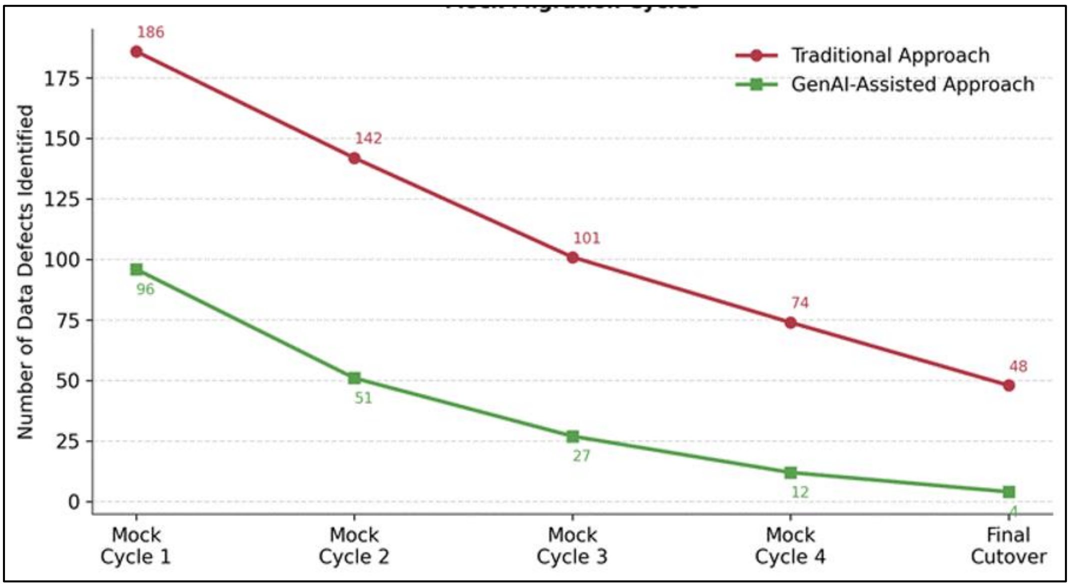

Figure 3: Defect reduction trend across successive mock migration cycles, traditional versus GenAI-assisted approach.

The GenAI-assisted approach resulted in just 4 residual defects after the final cutover rehearsal, while the traditional approach has 48, which is a 91.7% relative improvement, representing a significant improvement in data readiness for production cutover.

4.4 Module-Level Mapping Confidence

Table 3 shows the GenAI mapping confidence scores obtained in the schema mapping step for the five SAP modules examined in the case study.

Table 3: GenAI Field-Mapping Confidence by SAP Module

SAP Module	Legacy Object Count	GenAI Mapping Confidence (Avg.)	Manual Review Required
Finance (FI/CO)	18,420 records	94.2%	Low
Materials Management (MM)	26,150 records	89.7%	Moderate
Sales & Distribution (SD)	21,980 records	91.5%	Low
Human Capital Management (HCM)	9,340 records	83.1%	High
Plant Maintenance (PM)	7,610 records	87.6%	Moderate

The model was most confident with mapping structures for Finance (FI/CO) and Sales and Distribution (SD), where the legacy structures were more similar to the required mapping structures in S/4HANA and where enough historical experiences were available for the model. The group with the least confidence (83.1%) and the highest manual review burden was Human Capital Management (HCM) and that is due to the higher degree of structural heterogeneity, localization-specific fields, as well as privacyary attributes that need conservative human validation.

5 DISCUSSION

As highlighted in the findings of this case study, Generative AI has the potential to add value to enterprise data migration workflows in SAP transformation programs, especially in pattern recognition, repetitive mapping and documentation generation. The decrease of 53.1% in total effort and 65.5% in average defects are in line with the literature, which indicates that AI-based automation of processes creates most value when applied to well-structured, and high-volume tasks (Wamba-Taguimdje et al., 2020; Chui et al., 2018).

One thing to note: the results also reveal that there is still a need for human oversight, especially in governance sensitive areas like HCM data – where there are regulatory, privacy and localization factors that compromise the efficacy of purely automated mapping. This aligns with frames of responsible AI, which stress accountability and the importance of designing AI systems with a human in the loop in high-stakes operational settings (Ghallab, 2019; Sun & Medaglia, 2019).

In a pragmatic sense, the findings indicate that enterprise architects and SAP implementation partners (SAP-IPs) should think in tiers when integrating GenAI into the SAP system: highly automated GenAI integration for high-confidence modules that offer greater structural alignment, like Finance and Sales and Distribution, and adequate manual checks for modules that have higher legal, cultural, or structural demands, such as Human Resources, Customer Relationship Management, and Supply Chain Management. This tiered approach may help organizations realize efficiency gains without introducing undue governance risk.

There are also limitations to the study. The quantitative findings of this single case study are not guaranteed to be applicable to all enterprise scenarios and depend on data quality, maturity of the organisations and specific GenAI tooling. Furthermore, effort and defect measures were captured but long-term activity data quality results post cutover were not directly measured and should be the focus of future longitudinal studies.

6 CONCLUSION

This paper is an empirical case study on implementing Generative AI in a mock-up migration scenario for SAP ECC to SAP S/4HANA scenario. The automated framework, with automated profiling, LLM-assisted mapping, automated transformation rule generation and reconciliation reporting under ongoing human supervision, enabled us to see significant improvements in our effort, defect rates and accuracy in mapping in comparison to a traditional baseline. The results align with the idea that GenAI is best used as a layer that augments enterprise data migration, freeing up time for IT staff to handle more complex and governance-driven decisions. Future studies should be expanded to conduct analysis of several real-world SAP transformation programs, explore in-depth genetic data quality assessment post-cutover, and look into organization change management aspects playing a role in GenAI adoption in enterprise IT transformation.

REFERENCES

[1] Bansal, G., Chamola, V., Hussain, A., Guizani, M., & Niyato, D. (2022). Transformative effects of ChatGPT on modern education: Emerging era of AI chatbots. Internet of Things and Cyber-Physical Systems, 2, 129-140. https://doi.org/10.1016/j.iotcps.2023.06.002

[2] Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

[3] Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., ... Amodei, D. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems, 33, 1877-1901.

[4] Cao, L. (2022). AI in finance: Challenges, techniques, and opportunities. ACM Computing Surveys, 55(3), 1-38. https://doi.org/10.1145/3502289

[5] Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H. P. de O., Kaplan, J., ... Zaremba, W. (2021). Evaluating large language models trained on code. arXiv preprint arXiv:2107.03374.

[6] Chui, M., Manyika, J., & Miremadi, M. (2018). What AI can and can't do (yet) for your business. McKinsey Quarterly, 1, 97-108.

[7] Dilmegani, C. (2022). Data migration: Definition, best practices & tools in 2022. AIMultiple Research.

[8] Ghallab, M. (2019). Responsible AI: Requirements and challenges. AI Perspectives, 1(1), 1-7. https://doi.org/10.1186/s42467-019-0003-z

[9] Ghosh, A., & Chakraborty, D. (2022). Application of artificial intelligence in business analytics: A systematic review. International Journal of Information Management Data Insights, 2(2), 100109. https://doi.org/10.1016/j.jjimei.2022.100109

[10] Hasselbring, W., & Steinacker, G. (2017). Microservice architectures for scalability, agility and reliability in e-commerce. In 2017 IEEE International Conference on Software Architecture Workshops (ICSAW) (pp. 243-246). IEEE. https://doi.org/10.1109/ICSAW.2017.11

[11] Kalluri, R. (2020). Enterprise resource planning (ERP) migration strategies and challenges. International Journal of Computer Applications, 176(15), 1-6.

[12] Kumar, V., & Garg, M. L. (2021). Big data analytics and enterprise information systems: A review. Materials Today: Proceedings, 37, 3079-3082. https://doi.org/10.1016/j.matpr.2020.09.007

[13] Kupriyanovsky, V., Klimov, A., Chernykh, D., & Dobrynin, A. (2020). SAP S/4HANA and digital transformation of enterprise architecture. International Journal of Open Information Technologies, 8(6), 92-99.

[14] Laumer, S., Maier, C., & Weitzel, T. (2017). Information quality, user satisfaction, and the manifestation of workarounds: A qualitative and quantitative study of enterprise content management system users. European Journal of Information Systems, 26(4), 333-360. https://doi.org/10.1057/s41303-016-0029-7

[15] Ni, J., Young, T., Pandelea, V., Xue, F., & Cambria, E. (2022). Recent advances in deep learning based dialogue systems: A systematic survey. Artificial Intelligence Review, 56, 3055-3155. https://doi.org/10.1007/s10462-022-10248-8

[16] Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. Nature Medicine, 28(1), 31-38. https://doi.org/10.1038/s41591-021-01614-0

[17] SAP SE. (2021). SAP S/4HANA migration cockpit: Technical documentation and best practices. SAP Help Portal.

[18] Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of artificial intelligence in the public sector: Evidence from public healthcare. Government Information Quarterly, 36(2), 368-383. https://doi.org/10.1016/j.giq.2018.09.008

[19] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30, 5998-6008.

[20] Wamba-Taguimdje, S. L., Wamba, S. F., Kamdjoug, J. R. K., & Wanko, C. E. T. (2020). Influence of artificial intelligence (AI) on firm performance: The business value of AI-based transformation projects. Business Process Management Journal, 26(7), 1893-1924. https://doi.org/10.1108/BPMJ-10-2019-0411