



Digital Validation – Future Advances

May 2025

**A White Paper by the ISPE Commissioning & Qualification
Community of Practice**

Abstract

The life science validation space has been refined over the last few decades focusing on emphasizing risk-based learnings and methodologies. More recently the industry is witnessing a significant technological convergence unlike any other in recent times. As the vast amount of digital data generated over the years by industry is now becoming accessible and can be easily retrieved using various queries and tools.

In this new era where data is being recognized as an asset in the industry, it is accelerating innovation and advancements at a pace unprecedented until now. The life sciences industry has witnessed a steady increase in the adoption of Digital Validation Tools (DVTs) over the past decade. This trend has been propelled by a heightened emphasis on data integrity, the emergence of Pharma 4.0, Validation 4.0, and the necessity for accessible data on the validated states of systems and processes. These advancements aim to streamline monitoring and maintenance activities and enhance the understanding of processes and systems, via the adoption of iterative Quality Risk Management (QRM) principles in accordance with ICH Q9 to enable real time view of a process or system's validated status.

Acknowledgements

This White Paper represents the outcome of work done by the members of the Commissioning & Qualification Community of Practice Digital Validation Sub Committee as well as experiences and input from the individuals listed and does not reflect the views of any one individual or company.

Document Authors

Dave O'Connor	No deviation	Ireland
Phillp Jarvis	Grifols	Ireland
Matthew McMenamin	The Sentinel Consulting Group LLC	USA
John Cheshire	Headway Quality Evolution	United Kingdom
Subhabrata (Shubho) Purakayastha	Incyte Corporation	USA
Brandi M. Stockton	The Triality Group LLC	USA

Table of Contents

1	Introduction	4
2	Artificial Intelligence and Large Language Models	4
2.1	Large Language Models	4
2.2	AI/ML System Compliance and Validation	5
3	Machine Learning Opportunities and Data Standardization	6
3.1	ML Techniques for Enhancing Technical Documentation Review and the Importance of Human Oversight	6
3.2	Data Governance, Standardization, and the Role of DVT Vendors in Digital Transformation and Risk Management	6
4	Integration of DVTs and the Greater Digital Ecosystem	7
4.1	Current State, Limitations, and Challenges of Integrating DVTs	7
4.2	Data Stream Mapping, Dismantling Silos, and the Concept of an Industrial Cloud Platform Powered by Generative AI	7
4.3	Unified Namespace and Its Role in Data Integration and Standardization	7
4.4	Distributed Ledger Systems	8
4.5	Integration of DVTs with Maintenance, Calibration, and Periodic Validation Processes	8
4.6	Improved Execution Logics within DVTs	9
5	Conclusion	10
6	Abbreviations	11
7	References	11

Limitation of Liability

In no event shall ISPE or any of its affiliates, or the officers, directors, employees, members, or agents of each of them, or the authors, be liable for any damages of any kind, including without limitation any special, incidental, indirect, or consequential damages, whether or not advised of the possibility of such damages, and on any theory of liability whatsoever, arising out of or in connection with the use of this information.

© 2025 ISPE. All rights reserved.

All rights reserved. No part of this document may be reproduced or copied in any form or by any means – graphic, electronic, or mechanical, including photocopying, taping, or information storage and retrieval systems – without written permission of ISPE.

All trademarks used are acknowledged.

1 Introduction

Since 2020, the industry has seen a greater focus to transition away from traditional paper-based validation processes to a digital validation methodology. This shift initially embraced electronic validation repositories, enabling the digital upload, creation, and electronic approval of documentation, which was then securely stored within these systems. These platforms provided secure access and storage, along with the capability for electronic review and approval, aligning with electronic records and signature standards. However, these systems had limitations, notably in executing tests directly within them or managing document data, which left data integrity and process efficiency challenges associated with paper use unresolved.

Originally, many of these systems were adapted Document Management Systems (DMS) or custom-coded solutions that failed to fully address the comprehensive needs of the validation lifecycle. In the last five years, there’s been a significant shift towards investing in commercially available, configurable and customizable off-the-shelf solutions designed specifically for managing validation data digitally—from creation and execution to digital oversight.

These Digital Validation Tools (DVTs) facilitate digital execution that meets data integrity standards and allows for the management and sharing of validation data, thereby enhancing efficiency and compliance in ways that paper-based processes could not. This evolution marks a departure from merely digitizing paper-based processes (“paper-on-glass” approach) towards leveraging DVTs for substantial returns on investment and compliance benefits.

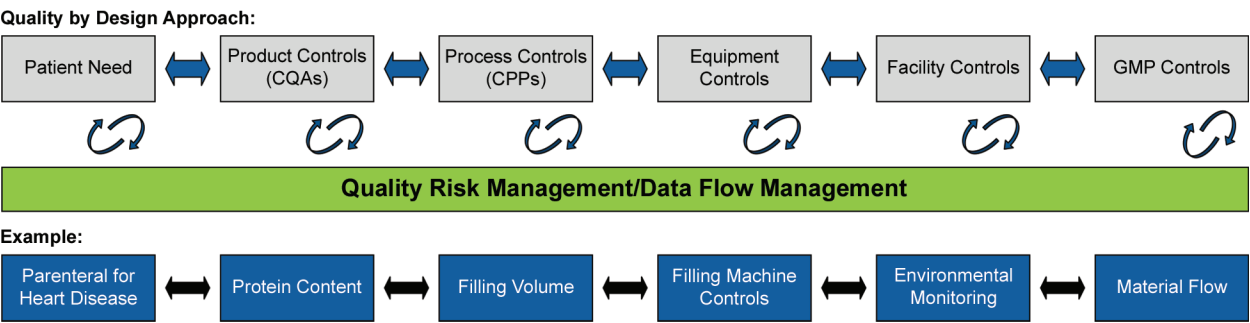
2 Artificial Intelligence and Large Language Models

2.1 Large Language Models

The implementation of Validation 4.0 marks an initial step towards achieving the principle of real-time verification. To fully enable real-time verification, it is important to integrate emerging technologies such as Large Language Models (LLMs) and Machine Learning (ML). Many organizations are starting to deploy LLMs within their private infrastructure to ensure compliance with data privacy and security standards.

To explore how to achieve Validation 4.0, consider Figure 1, which outlines the principles of the Quality by Design (QbD) approach, QRM and data flow management:

Figure 2.1: Principles of Quality by Design



In Figure 2.1, data flow from the manufacturing process interacts with both the control strategy and the QRM system, enabling real-time verification of changes within the manufacturing process.

Take an example of an equipment control change in the validated process related to a Critical Process Parameter (CPP) such as temperature control in a bioreactor.

In a traditional paper-based approach, a change control would be initiated to revise the User Requirement Specification (URS) and then a test document would be prepared to verify that the change in temperature control (i.e., the addition of a new alarm) works as intended. This approach is static, time consuming, and requires extensive preparation, review, approval and execution.

However, if the technical specifications and documents are stored in a DVT, machine learning could potentially detect the change. LLMs and Artificial Intelligence (AI) could then generate the necessary test verification based on the documentation changes.

Via API interfacing with the actual temperature control system (e.g., a Distributed Control System (DCS) system) the verification code could be executed within the DVT, automatically incorporating the alarm verification into the test document.

This combination of emerging technologies in a digital ecosystem could transform a static paper-based process into a nearly real-time verification system, where AI/LLM generated outputs are reviewed and approved on the tool by the appropriate stakeholders as a quality control measure, both pre and post execution.

These tools can play a significant role in enabling real-time verification and streamlining validation processes. By leveraging the power of natural language processing and machine learning, these technologies can analyze vast amounts of data, identify patterns, and generate insights that can help optimize validation activities.

LLMs and generative AI are particularly effective at processing large volumes of information and generating documentation based on the understanding of these documents. This has several important applications in validation, such as the generation of technical content for test verification and the development of user requirements based on the latest regulatory guidance and standards. However, it is important to emphasize the need for human oversight to ensure the accuracy and compliance of the generated content.

2.2 AI/ML System Compliance and Validation

It is essential to consider factors such as data quality, algorithmic stability, model performance, model explainability, and documentation to ensure the reliability and effectiveness of AI/ML models.

For most systems incorporating AI/ML, many elements of the traditional computerized system lifecycle and compliance/validation framework remain applicable. These include aspects related to specification and verification of user interfaces, reporting, security, access controls, and data lifecycle management.

Appendix D11 of *ISPE GAMP® 5 (Second Edition)* [1], describes an approach and lifecycle framework for the use of ML within the life sciences industry and GxP regulated environments.

3 Machine Learning Opportunities and Data Standardization

3.1 ML Techniques for Enhancing Technical Documentation Review and the Importance of Human Oversight

ML techniques can be employed to enhance the speed and quality of creating and reviewing technical documentation in the pharmaceutical industry. For example, a typical CQV engineer often spends a significant amount of time generating and reviewing documents such as material certifications. ML algorithms trained on Optical Character Recognition (OCR) can convert scanned documents into machine-readable text. Text parsing and Named Entity Recognition (NER) models can then extract relevant information such as material type, batch number, specifications, and testing results. Anomaly detection algorithms can help identify discrepancies or missing data, ensuring the authenticity and accuracy of the documents. The extracted data can be subsequently fed directly into a database or Material Resource Planning (MRP) system, preventing data duplication and ensuring alignment with existing data structures.

While ML techniques can greatly accelerate and improve the review process, it is crucial to emphasize the importance of human oversight in these semi-autonomous applications. Final approval should remain with human experts to ensure accuracy and compliance of records with regulatory standards.

3.2 Data Governance, Standardization, and the Role of DVT Vendors in Digital Transformation and Risk Management

Effective data governance [2] and standardization are critical for successful digital transformation and risk management in the pharmaceutical industry. As companies progress towards Industry 4.0, the focus shifts towards data-driven validation processes. To maximize the value of advanced DVTs and enable real-time testing, it is essential to develop a standardized data framework across global operations. This framework serves as the foundation for data flow and integration, allowing validation protocols to directly pull data directly from endpoints such as manufacturing equipment and quality control systems.

ICH Q9(R1) [3] highlights the importance of applying QRM principles to the design, validation, and technology transfer of advanced production processes, analytical methods, data analysis methods, and computerized systems. To ensure that recent digital tools do not introduce uncontrolled risks, effective data governance must be in place and comprehensive data structure framework should be established by digitization software providers.

DVT vendors play a vital role in this transformation by enhancing their offerings and making their data more accessible to new market entrants. This fosters innovation and competition, ultimately benefiting customers through improved products and services. As DVT vendors continue to evolve, their tools pave the way towards achieving the goals of Validation 4.0, where data integrity by design becomes feasible. Through integration with other systems or utilizing multiple tools, it is now possible to capture and make visible data from various endpoints across the entire validation process lifecycle.

Data standardization is the critical enabler for real-time validation and seamless data flow integration in the pharmaceutical industry. By developing a standardized data framework across global operations, companies can leverage advanced DVTs to perform real-time testing and draw data directly from endpoints such as manufacturing equipment and quality control systems. This standardization allows for the harmonization of data formats and facilitates the storage and access of data across various systems and applications.

Standardizing the creation and structure of data enables seamless querying and analysis, empowering organizations to gain valuable insights and make informed decisions by leveraging data from diverse sources. To fully capitalize on the benefits of advanced tools and technologies in streamlining data management, the industry must embrace a paradigm shift in how information is organized and structured. The relational database model, coupled with robust data governance, has emerged as a popular solution for efficiently storing and managing complex data sets.

4 Integration of DVTs and the Greater Digital Ecosystem

4.1 Current State, Limitations, and Challenges of Integrating DVTs

Many current DVTs, whether deployed within organizations or commercially available, are still in a transitional phase. While they represent an improvement over traditional paper-based and “paper-on-glass” methods, these systems often do not fully embrace a total digital transformation paradigm shift. DVTs frequently replicate rigid paper-based processes in a digital format, including documents like Installation Qualification (IQ), Operational Qualification (OQ), and Requirement Specifications (RS), whilst, on the other hand, also attempting to adopt more modern approaches such as Computer Software Assurance (CSA) and Agile methodologies.

These digital systems often retain structures inherited from their paper-based predecessors, often as a result of digital validation providers adapting existing paper processes to meet client needs. Although DVTs incorporate basic linkages and workflows, the data within these systems remains largely static, requiring extraction for analysis or trending. The integration capabilities of these systems are limited, with few APIs for data sharing and a lack of unified access protocols across various platforms, including DVTs, DMS, Enterprise Information Management Systems (EIMS), and Asset Management Systems (AMS).

Additionally, change management processes in most DVTs, despite being designed for digital operations, still require manual intervention for updates to individual components such as RS or OQ documents. Some systems focus narrowly on specific areas such as computer system validation or equipment qualification, without addressing the broader spectrum of validation needs.

4.2 Data Stream Mapping, Dismantling Silos, and the Concept of an Industrial Cloud Platform Powered by Generative AI

To overcome the challenges of integrating DVTs and create a cohesive digital ecosystem, organizations must prioritize comprehensive data stream mapping. This process involves considering all existing systems, their versions, data access capabilities, and the flow of data between them. Data stream mapping is a crucial preparatory step in the development of an integrated digital ecosystem and dismantling operational silos.

The concept of an industrial cloud platform powered by generative AI offers a promising solution to these integration challenges. Such a platform would provide industry-relevant, direct business outcomes by packaging Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) around a robust data integration and contextualization backbone. This backbone, often referred to as an industry data fabric, would be complemented by libraries of composable business capabilities, including use case templates, domain specific data models, pre-built analytics schemas and dashboards, and industry-specific generative AI tools.

An industrial cloud platform powered by generative AI would provide an end-to-end, integrated set of tools to streamline and accelerate application development across all data use cases. It would reduce the cost of implementation, integration, and maintenance while ensuring common governance, metadata management, and unified access management.

4.3 Unified Namespace and Its Role in Data Integration and Standardization

A Unified Namespace (UNS) plays a vital role in data integration and standardization within the pharmaceutical industry. UNS provides a consistent and standardized way to identify, access, and manage data across various systems and platforms. By establishing a common naming convention and hierarchical structure, UNS enables seamless data integration and interoperability, reducing the complexity of managing data across disparate sources.

In the context of DVTs, UNS facilitates the integration of validation data with other relevant data sources, such as maintenance records, calibration data, and process control systems. By standardizing the naming and structure of data elements, UNS allows for the creation of a unified data model that can be leveraged by various applications and systems involved in the validation process.

UNS also supports data governance efforts by providing a clear and consistent framework for organizing and managing data assets. It helps in establishing data ownership, defining access controls, and ensuring data quality and integrity across the organization. By promoting data standardization through UNS, companies can enable real-time validation, streamline data flow integration, and facilitate the implementation of advanced analytics and AI-powered solutions.

4.4 Distributed Ledger Systems

Distributed Ledger Systems (DLS), such as blockchain, are increasingly being explored for use in regulated environments as supporting GxP systems, although their current application scope is limited. These systems offer a revolutionary approach to digital validation in software development by providing an immutable, transparent, and decentralized record of transactions. As familiarity and availability of such technology for GxP increases, its use in regulated environments is expected to grow. As digital validation data is increasingly originating from segregated processes from various sources (both internal and external), the use of a decentralized database aids in the accessibility and ability to leverage and interpret such data.

In the context of software validation, DLS can be used to ensure the authenticity of code, track version histories, and verify the integrity of updates or patches. For regulated companies, these are all assurances needed from a DVT system to be suitable to support GxP conduct; therefore, the use of DLS is becoming a promising solution for future-ready DVTs. By leveraging blockchain's tamper-proof nature, organizations can maintain integrity of their validation data and trust in the development lifecycle, reducing the risk of fraud and errors while enhancing accountability and transparency.

For further information, refer to *ISPE GAMP® 5* (Second Edition) [1], Appendix D10 – Distributed Ledger Systems (Blockchain).

4.5 Integration of DVTs with Maintenance, Calibration, and Periodic Validation Processes

Integrating DVTs with maintenance, calibration, and periodic validation processes is essential for creating a holistic and efficient validation ecosystem. Traditionally, organizations often siloed these activities into separate departments, leading to the development of discrete methodologies that do not immediately interact, despite the common goal of ensuring product quality.

To overcome this challenge, organizations should adopt a more integrated approach to process design, allowing for the harmonization of data formats, and the standardization of data storage and access across various validation related activities. This approach is already well-established in DVTs, which allow the integration of existing “paper-on-glass” templates to capture digital testing data.

Integrating DVTs with maintenance and calibration processes enables organizations to leverage validation data captured during the Commissioning, Qualification, and Validation (CQV) phase, eliminating the need for redundant verification processes in separate digital tools. This integration can be facilitated using APIs, which enables seamless data exchange between DVTs and other client databases.

Furthermore, integrating DVTs with periodic validation processes allows for the continuous assessment and maintenance of the validated state of computerized systems and monitoring of associated processes. By collecting data from existing quality processes, such as system suitability testing performance, hardware calibration and maintenance, software incidents, and change control, organizations can leverage DVTs and UNS to capture process data in a harmonized format and location. This approach enables more effective system management and provides easier access to real-time data on system performance, which can be used to assess the suitability of a system and reduce the effort required for revalidation when necessary.

Integration of DVT can therefore foster collaboration and harmonization between organizationally segregated departments and drive a greater awareness of information and its impact as it occurs.

System periodic review is often an isolated activity to review the status of a validated system. DVT alone may streamline the process; however, this requires integration with data from existing processes to fully optimize the approach. Integration with a Corrective Action and Preventive Action (CAPA) and deviation management system can automate a review of incidents and changes made to the system during the period that can in turn identify impact assessment through risk management system, Electronic Laboratory Notebook (ELN), or LIMS integration, for example. Integration of DVT for periodic validation review can reduce the requirement for an isolated activity altogether, adapting the process to a more real-time and real-data dashboard review of a system status.

It is recommended for organizations to focus on data-centric validation rather than documentation-driven approaches. There is a misconception that utilizing DVT ensures an agile and efficient approach to validation; however, without further critical thinking, risk assessment, and process mapping this is unlikely to be met.

4.6 Improved Execution Logics within DVTs

There is very limited logic behind form filling that ensures all required data is correctly entered before submission. This results in incomplete data entry (blanks), or data not entered within the acceptable range (out of specification) during the execution phase. These errors often go unnoticed until the review process, which leads to time consuming rework and exception handling.

To address this deficiency, DVTs should incorporate a combination of validation techniques, user interface design principles, and programming logic to reduce human error during execution, and accelerate document turnaround times. Below are some basic rules of form filling used in other industries that could be transposed to the DVT space:

- **Validation Rules**

Ensure each input field in the form is associated with specific validation rules, such as:

Required Fields: Fields that must be filled out before submission

Input Format: Fields can enforce specific formats (e.g., number, text, file format)

Data Type Validation: Fields can be checked to ensure that the type of data entered is valid (e.g., numbers in a numeric field)

- **Real-Time Validation**

Forms can validate user input in real-time. This can involve:

Immediate Feedback: As users fill out each field, they receive instant feedback about whether their input is valid. For example, if a user enters an out of specification value, the system highlights the results immediately.

Visual Cues: Fields may change color (e.g., from red to green) or display checkmarks to indicate valid input, helping users identify errors quickly.

- Submission Logic

Before the form can be submitted, a final validation step is performed. This includes:

Overall Validation Check: When the user attempts to submit the form, the system checks all fields against the validation rules. If any fields are invalid or empty, the form submission is halted.

Summary of Errors: If the form cannot be submitted, a summary of error messages or warnings is displayed to guide users on correcting their input.

- User Interface Design

Good user interface design also plays a crucial role in guiding users through the form filling process:

Logical Flow: Forms are organized in a logical order, grouping related fields together and following a progression that feels intuitive to users.

Disabled Reviewed by Button: The button can remain disabled until all validation rules are satisfied, signaling to the user that they cannot proceed until all required fields are correct.

5 Conclusion

This concept paper explored the future advances in DVTs and their impact on the pharmaceutical industry. Key points to consider for future focus on DVTs include:

1. The need for easily accessible data with minimal queries to extract data from DVTs.
2. The role of AI, LLMs, and ML in enabling real-time verification, streamlining validation processes, and enhancing the review of technical documentation.
3. The importance of data governance, standardization, and the role of DVT vendors in digital transformation.
4. The challenges and current limitations of integrating DVTs and the need for a shift in mindset towards full digital transformation.
5. The future potential of technological tools such as blockchain, predictive analytics, and continuous validation in shaping the future of DVTs and the pharmaceutical industry.
6. Importance of verification of the data used for the development of these AI tools to ensure that they are adequate for generating a certain output and are not biased or corrupted. The quality of an AI/ML application depends on the quality of the data, which will form the framework for validation in the near future.

As the pharmaceutical industry continues to evolve and embrace digital transformation, DVTs will become increasingly crucial for enabling efficient, compliant, and data-driven validation processes. The integration of AI, LLMs, and ML technologies will further enhance the capabilities of DVTs, allowing for real-time verification, streamlined documentation review, and improved risk management.

The adoption of industry-wide standardization and the development of open ecosystems will facilitate a better integration and control of DVTs, enabling seamless data flow and collaboration across the pharmaceutical value chain.

As DVTs continue to advance and mature, they will become an integral part of the pharmaceutical industry's digital transformation journey, driving innovation, efficiency, and compliance.

6 Abbreviations

AI	Artificial Intelligence
AMS	Asset Management System
CAPA	Corrective Action and Preventive Action
CPP	Critical Process Parameter
CQV	Commissioning, Qualification, and Validation
CSA	Computer Software Assurance
DCS	Distributed Control System
DLS	Distributed Ledger System
DMS	Document Management System
DVT	Digital Validation Tool
EIMS	Enterprise Information Management System
ELN	Electronic Laboratory Notebook
IaaS	Infrastructure as a Service
IQ	Installation Qualification
LLM	Large Language Model
ML	Machine Learning
MRP	Material Resource Planning
NER	Named Entity Recognition
OCR	Optical Character Recognition
OQ	Operational Qualification
PaaS	Platform as a Service
QbD	Quality by Design
QRM	Quality Risk Management
RS	Requirement Specification
SaaS	Software as a Service
UNS	Unified Namespace
URS	User Requirement Specification

7 References

1. *ISPE GAMP® 5: A Risk-Based Approach to Compliant GxP Computerized Systems*, International Society for Pharmaceutical Engineering (ISPE), Second Edition, July 2022, www.ispe.org.
2. *ISPE GAMP® RDI Good Practice Guide: Data Integrity by Design*, International Society for Pharmaceutical Engineering (ISPE), First Edition, October 2020, www.ispe.org.
3. International Council for Harmonisation (ICH), ICH Harmonised Tripartite Guideline, *Quality Risk Management – Q9(R1)*, Step 4, 18 January 2023, www.ich.org.

4. FDA Guidance for Industry: Data Integrity and Compliance with CGMP, Questions and Answers, December 2018, US Food and Drug Administration (FDA), www.fda.gov.
5. MHRA Guidance: 'GXP' Data Integrity Guidance and Definitions, Revision 1, March 2018, Medicines & Healthcare products Regulatory Agency (MHRA), www.gov.uk/government/organisations/medicines-andhealthcare-products-regulatory-agency.
6. General Principles of Software Validation; Final Guidance for Industry and FDA Staff, January 2002, US Food and Drug Administration (FDA), www.fda.gov.
7. FDA Draft Guidance for Industry: Computer Software Assurance for Production and Quality System Software, September 2022, US Food and Drug Administration (FDA), www.fda.gov.
8. WHO Technical Report Series, No. 966 – Annex 5: Guidance on Good Data and Record Management Practices, 30 September 2016, World Health Organization (WHO), <https://extranet.who.int/prequal/medicines/who-technical-report-series>.
9. PIC/S Guidance: PI 041-1 Good Practices for Data Management and Integrity in Regulated GMP/GDP Environments, 1 July 2021, Pharmaceutical Inspection Co-operation Scheme (PIC/S), www.picscheme.org/.
10. FDA Guidance for Industry: Process Validation – General Principles and Practices, January 2011, US Food and Drug Administration (FDA), www.fda.gov.



3001 N. Rocky Point Dr. E., Suite 200-242, Tampa, Florida 33607 USA
Tel: +1-813-960-2105, Fax: +1-813-264-2816

www.ISPE.org