AI / ML and Data Enablement in Generic Drug Manufacturing

Jeremy Evans, Ph.D.

Director AI / ML Innovation – Advanced Manufacturing Technology

June 8, 2026 - Generic Drug Science and Research Initiatives Public Workshop

Our enduring mission

> To improve global health through public standards and related programs that help ensure the quality, safety, and benefit of medicines, dietary supplements, and food ingredients.

USP portfolio: Beyond standards

- **Education for regulators and industry**
- **Supply chain and manufacturing services**
  - Advanced manufacturing
  - Medicine Supply Map
  - Pharmaceutical Product Quality Testing (PPQT)
  - Technical services
- **Verification services for ingredients**
- **R&D analytical solutions**
- **Digital solutions**

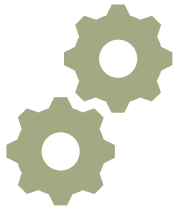
Opportunities Offered Through Data Driven Solutions

Data Quality, Integrity, Consistency

Avoid Manual Transcription Errors

Automated Data Quality Checks

Leverage Taxonomies and Controlled Processing



Operational Efficiency

Rapid Processing

Direct Data Integrations

Automate Overhead

Speed to Market

Avoiding Scheduling Delays

Process and Formulation Optimization

Faster Regulatory Submission

© USP

Applications of Digital Methods to Generic Drug Development

Applications and Enablement at USP

- **Medicine Supply Map**
  - Leverages millions of data points to predict drug shortages.
- **Digital Reference Standards (USP-ID)**
  - Automated identification and quantification through NMR
- **Compendial modernization and digitization**
- **Adapting pharmacopeial standards for digital and AI-enabled workflows**
- **Global Substance Registration System**
- **In Silico Modelling**

Industry Applications

- **Document Intelligence, Authoring, and Digitalization**
- **Quality Process Support**
- **Retrosynthesis and Route Scouting**
- **Chemometrics, Monitoring, and Assessment**
- **Process Modelling and Scale-Up**

© USP

Excipient Nomenclature and Innovation Activities

Nomenclature

- USP-NF nomenclature leveraged by USAN Council and GSRS for preferred non-proprietary naming.
- Expert Committee on Nomenclature and Labeling.
- Excipient Nomenclature Guideline Established 2022.

1. https://www.usp.org/health-quality-safety/compendial-nomenclature

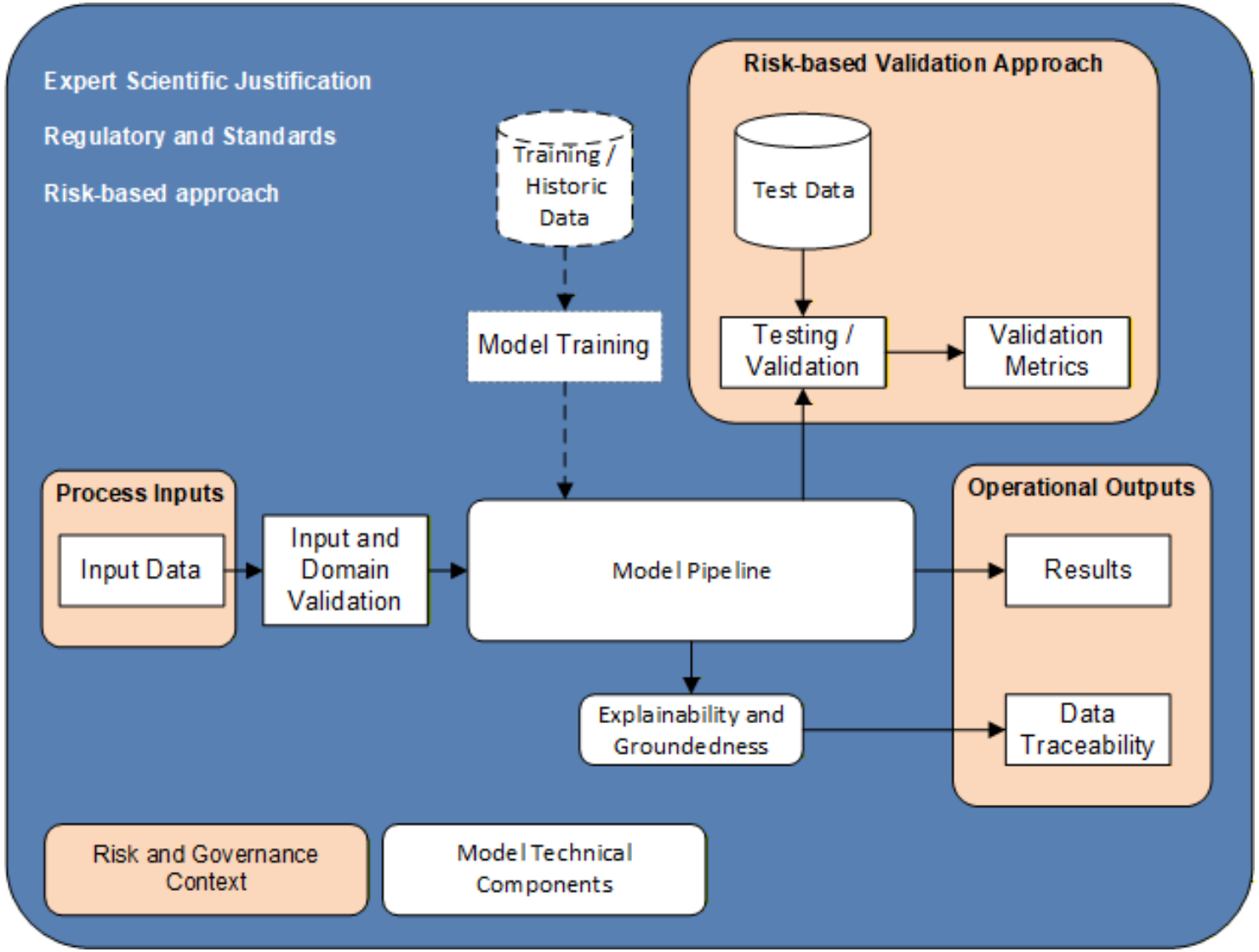
Excipient Quality and Innovation

- Fostering Novel Excipients in Drug Products
- Novel Excipients Knowledge Hub
- Stimuli article on Strategic Approaches to Excipient Toxicity Assessment – PF 52(3)
- Excipient Standards Catalog

2. https://www.usp.org/excipients

Anchored to Standards and Governance

- Digital tools operate within a context of scientific justification and accountability.
- Risk-based validation, monitoring, and governance approach.
  - Including both positive and negative validation.
- Outputs grounded in reliable data and standards relevant to inputs.



© USP

Data Challenges and Enablement Best Practices

Data Challenge	Mitigation Strategies
Unclear Data Provenance and Prior Transformations	• Clarify the intended data set use case(s) • Ensure fields are well defined and used consistently • Including consistent meaning of “Null” entries • Traceability to source • Disaggregate data where possible – include multiple statistical measures (max, min, quartiles, etc.) otherwise
Linking Challenge Across Data Sets	• Leverage controlled vocabularies and taxonomies for categorical attributes • Capture foreign keys to associated contextual databases (ex GSRS)
Biased and Censored Data Sets	• Data provenance details can highlight potential for bias • Beware of “positive” bias in published / submitted information
Manual Ingest and Cleaning Steps	• Programmatic interfaces / APIs • MCP to make data sets available to agents • Validated ingest pipelines

© USP

AI / ML Applications for IID and MDE

Advantages of Data Driven Techniques

- Automated extraction of input data from ANDAs or Labels.
- Direct calculation of Potency Amount and MDE from Inputs.
- Integrate substance data from GSRS.
- Present combined details and source information for human-in-the-loop review.
- Grouping and clustering IIDs.

Likely Challenges to Solve

- Validating reliable extraction across formats and document structures.
- Resolving naming and unit mismatches.
- Ensuring consistent calculations and definitions between FDA and industry.
- Human-in-the-loop user experience.
- Recognizing and handling edge cases.

© USP

Thank you

The standard of trust