From Data Inventory to AI-Ready Knowledge Architecture

Unlocking Scalable Intelligence, Compliance, and Accelerated R&D in Veterinary Medicines Development

Dr Cat Stirling

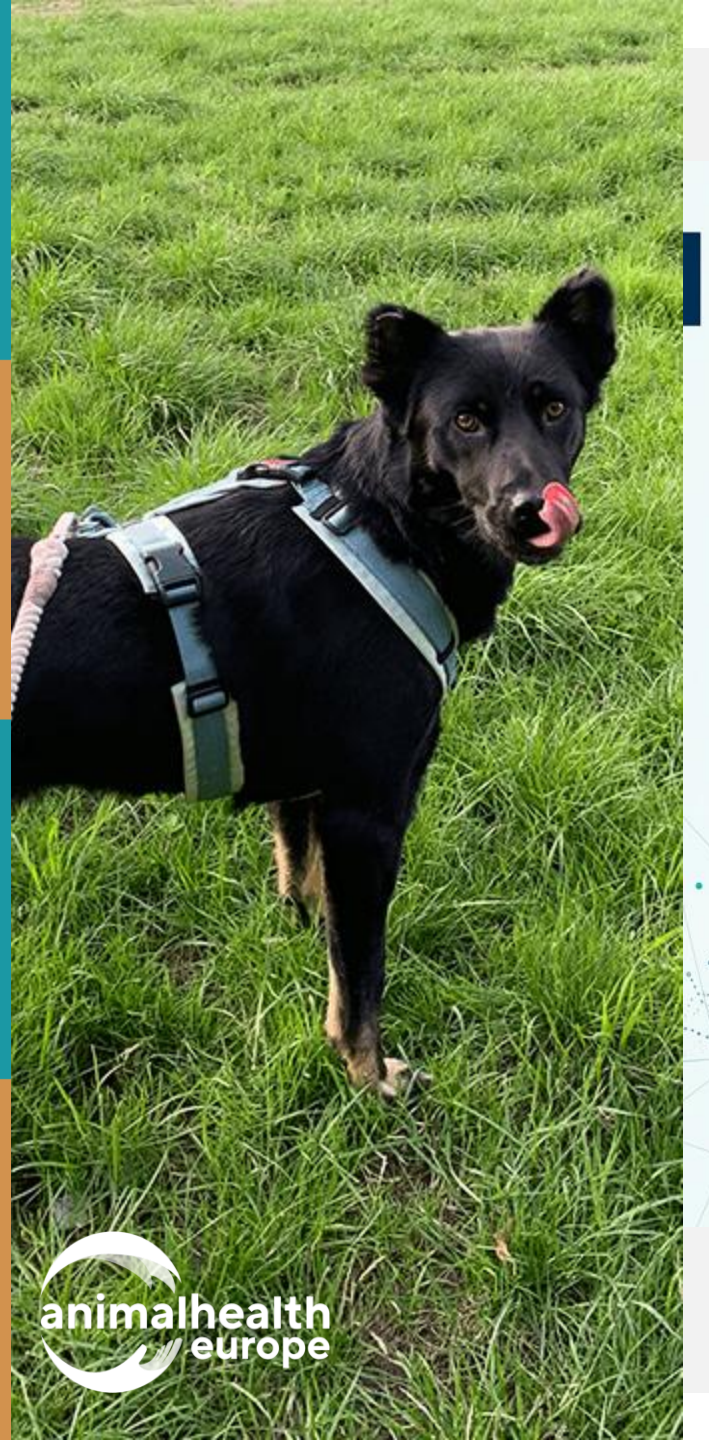
6th Veterinary Big Data forum

animalhealth europe
THE VOICE OF THE ANIMAL MEDICINES INDUSTRY



Disclaimer

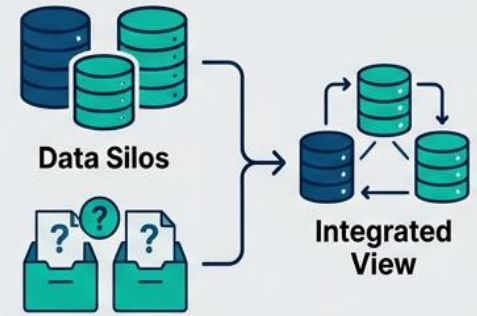
- The views expressed in this presentation are the personal views of the presenter. They shall not be understood or cited as opinions of Elanco Animal Health.
- The presenter does not have any conflict of interests.
- All Images were generated by AI, unless stated otherwise



The Core Challenge: Data-Rich, Knowledge-Poor

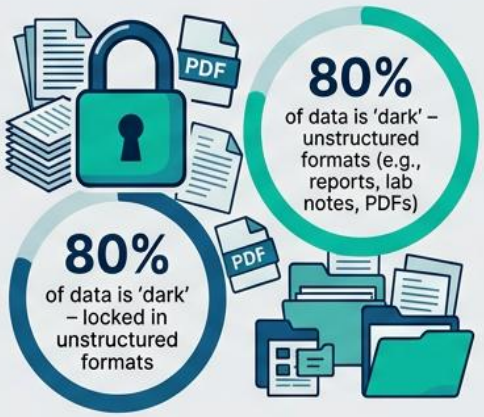
Having data is not the same as being AI-ready

The Data Inventory Paradigm



Disconnected datasets without context prevent meaningful synthesis.

The Dark Data Dilemma



Crucial insights are trapped and inaccessible to analysis.

The Inefficiency Tax



60-80%

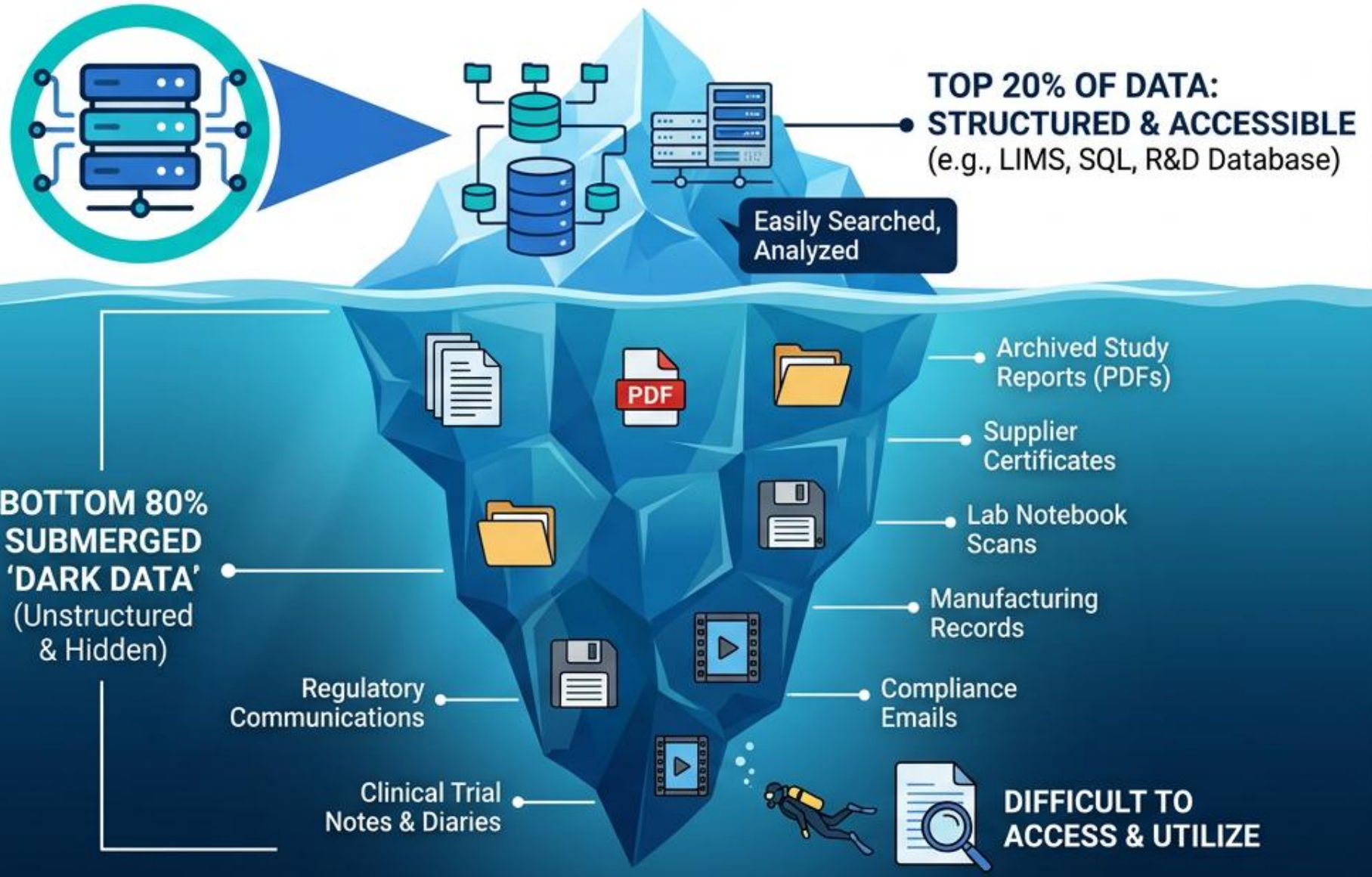
of researcher time is spent on data discovery and cleaning

Valuable time wasted searching for and preparing data instead of analyzing it.

AI

AI-Ready Data Foundation | Veterinary Pharma Insights

VETERINARY PHARMACEUTICAL DATA: REVEALING THE 'DARK DATA' ICEBERG





THE DARK DATA DILEMMA

Estimated that ~80% of corporate data remains unstructured and 'dark', leading to missed opportunities and inefficiencies.

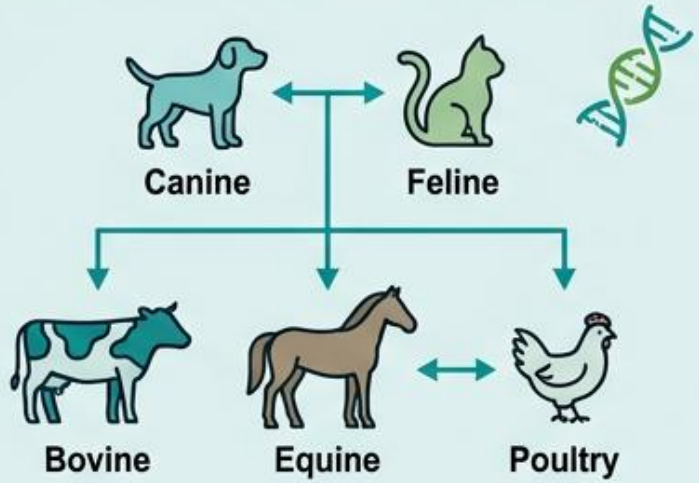


OPERATIONAL INEFFICIENCY TAX

Veterinary Pharma companies may face up to **60-80%** in operational inefficiencies when critical information is trapped or hard to find.

Veterinary Nuances vs. Human Pharma in Data-Driven R&D

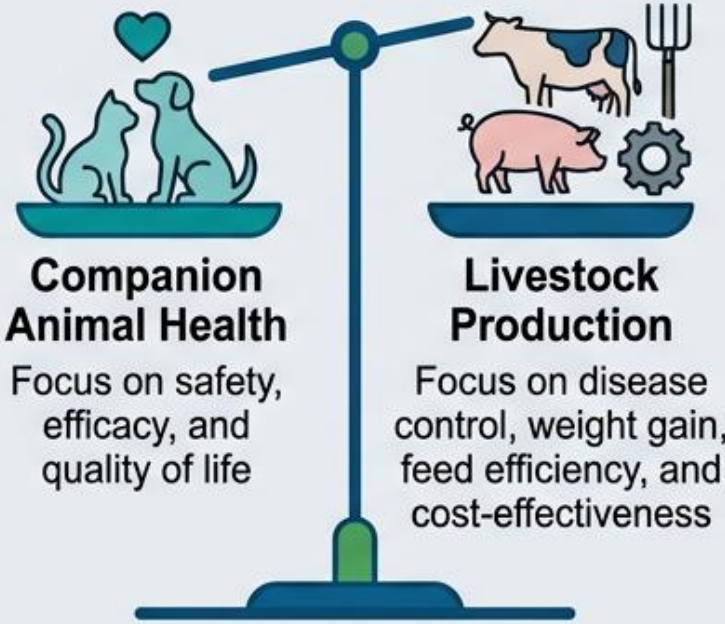
Interspecies Heterogeneity



Complexity from distinct biological systems, requiring tailored R&D approaches for each species

Multiple variations in metabolic pathways and drug responses across diverse animals

Dual Health & Production Endpoints



Companion Animal Health

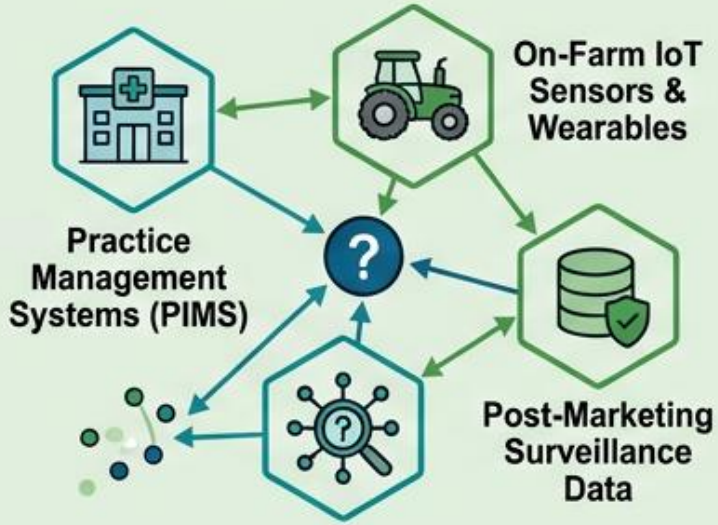
Focus on safety, efficacy, and quality of life

Livestock Production

Focus on disease control, weight gain, feed efficiency, and cost-effectiveness

Unique regulatory requirements: Maximum Residue Levels (MRLs) and Withdrawal Periods

Fragmented Real-World Data (RWD)



Disparate data sources, lack of standardisation, and data sharing challenges

Limited data integration hindering large-scale insights



The Paradigm Shift: Inventory vs. Architecture

Moving from passive catalogues to active, contextualised knowledge

Data Inventory (Legacy)

Static Files and Tables

Disparate Systems

Keyword Search

Finds words, not meaning

Brittle Prompt Queries

! Invalid

Lacks context for prompts

High Hallucinations

Fragment, incorrect factoids

HIGH HALLUCINATION RISK

AI-Ready Knowledge Architecture (Future)

Entities and Concepts

Drug Compound

Indication

Species

Regulatory Body

Connects facts as meaningful concepts

Semantic Reasoning

Logic

Deduces relationships

Understanding intent

Understanding intent

Universal FAIR Standards

F A I R

Universal FAIR Standards

GraphRAG and Dynamic Data Lineage

Graph database visualisation

Real-time Lineage

Provenance Tracking



The Foundational Layer: FAIR Principles & Ontologies

Standardising the language of animal health

FAIR
By Design

- Findable
- Accessible
- Interoperable
- Reusable

Controlled Vocabularies & Ontologies

VeDDRA
Pharmacovigilance

Uberon & SNOMED-CT
Anatomy & Disease

EMA SPOR / UPD standards
Regulatory IDs

Global Identifiers & Data Lineage

Persistent GUIDs/URIs connecting:

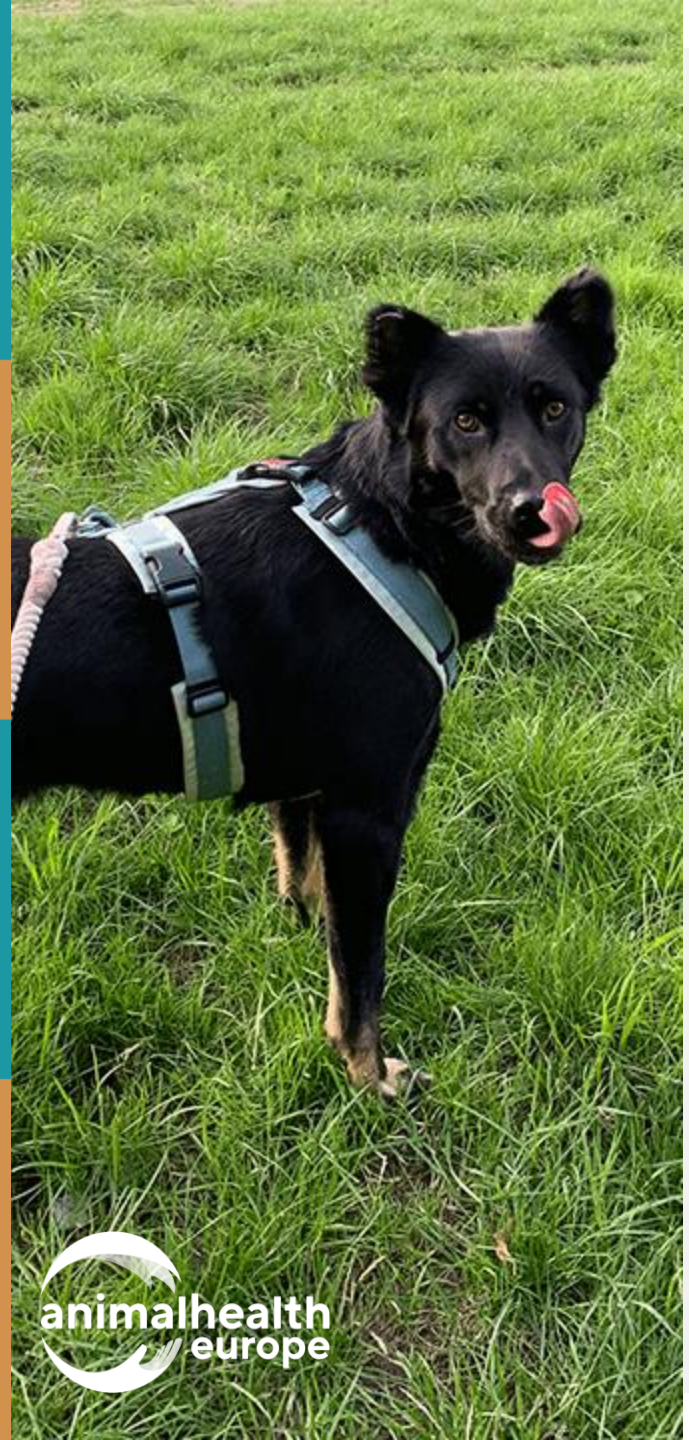
Active substances → Study protocols → Animal target species



Architectural Blueprint: The Semantic Knowledge Graph

Connecting dots across the product lifecycle.

- **Enterprise Knowledge Graph (EKG):** Triples of knowledge (e.g., [Compound X] → *inhibits* → [Pathway Y] → *expressed in* → [Canine Breed Z]).
- **Multi-Modal Ingestion:** Unifying structured data (bioassay tables, pharmacokinetic curves) with unstructured text (study reports, peer-reviewed literature, regulatory correspondence).
- **Graph-RAG (Retrieval-Augmented Generation):** Guiding LLMs with factual graph relationships to eliminate hallucinations in scientific queries



High-Value Use Case 1: Target Discovery & Interspecies Repurposing

Accelerating early pipeline eipeline exploration through cross-species translation

Translational Biology Conservation

DNA sequence

DNACACGAATACAATCCCACAG
DNACACCAAAACATTCCCACCG

Model organisms

Protein — Protein 1 — Conserved domain

Protein — Protein 2 — Conserved domain

Protein — Protein n — Conserved domain

Historical Assay Mining & Toxicity Prediction

Massive amounts of historical bioassay data

AI / Machine Learning

Pre-screening for toxicity

In-vivo study

Multi-Indication Screening

Single compound molecule testing tested simultaneously across different

Livestock health

Companion animal health



High-Value Use Case 2: Smarter Studies & the 3Rs

Optimising study design to drive **R**eplacement, Reduction, and **Refinement**

Synthetic Control Arms



Historical datasets used as a digital control group

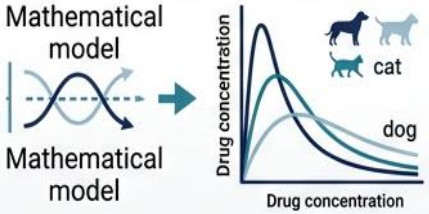
Animal trial group

Reduces number of live animals required

Minimises standard treatment arms using **real-world data**

Reduction

Predictive Dose Formulation



Matches formulation kinetics with species physiological models

Enables early screening of compounds

Refinement

Protocol Risk & Recruitment Optimisation



Identifies potential trial bottlenecks and recruitment challenges

Flags study failure risks in advance

Replacement/ Refinement

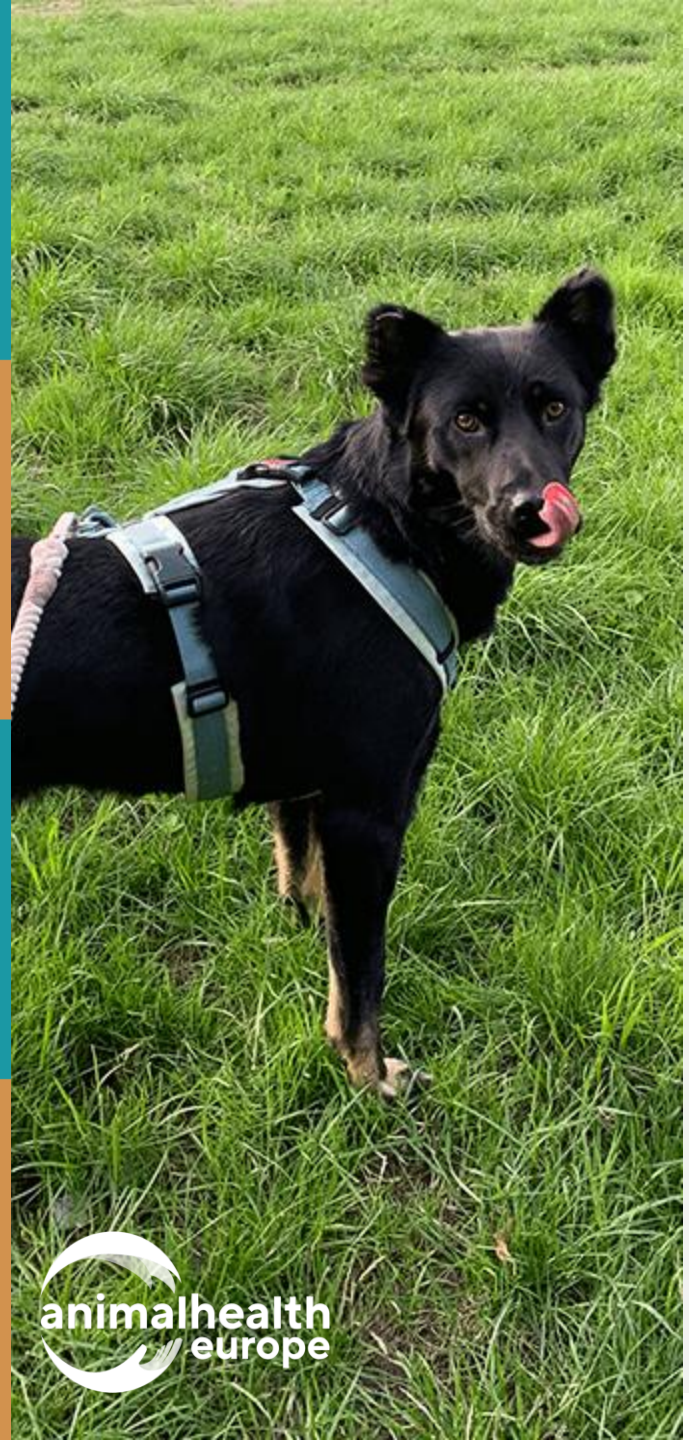
BioPharma Solutions | VetAnalytics



High-Value Use Case 3: Regulatory Intelligence & Dossier Automation

Streamlining global submissions and post-market safety.

- **Dossier Synthesis:** Extracting, cross-referencing, and drafting Module summaries (Parts I–IV / CTD equivalent) from primary study archives.
- **Global Variation Impact Analysis:** Automatically evaluating how a CMC or manufacturing change affects multi-region registrations.
- **Real-Time Pharmacovigilance & Signal Detection:** Continuous semantic ingestion of adverse events linked directly to substance registries and batch histories.



Trust, Governance & Regulatory Compliance (GxP)

AI readiness is meaningless without compliance and auditability.

- **Data Lineage & Provenance:** Full traceability—knowing precisely which study report, table, or calculation informed an AI-generated insight.
- **Human-in-the-Loop (HITL):** Embedding mandatory scientific and regulatory validation gates for all AI-assisted outputs.
- **Regulatory Alignment:** Complying with emerging AI guidelines from global regulatory bodies (EMA Big Data / AI Workplan, FDA guidelines).



Strategic Implementation Roadmap

A phased transformation journey

- **This is a big strategic investment for industry - example**

  - **Phase 1: Foundations (Months 1-6):** Audit key data silos; standardise core taxonomies (VeDDRA, species); establish metadata conventions.
  - **Phase 2: Semantic Pilot (Months 6-12):** Build a domain-specific Knowledge Graph pilot (e.g., focused on a single therapeutic area or Pharmacovigilance).
  - **Phase 3: Scale & GraphRAG (Months 12-18):** Connect regulatory dossiers, clinical archives, and integrate AI agentic workflows.
  - **Phase 4: Enterprise Intelligence (Months 18+):** Enterprise-wide automated insights, predictive trial modelling, and continuous governance.



Summary & Conclusions

Knowledge architecture is a strategic capability, not just an IT project.

- **Key Takeaways:**
  - Data storage alone does not create and drive medicines innovation; semantic connectivity and AI-readiness do.
  - Addressing veterinary-specific multi-species complexities early prevents costly downstream technical debt.
  - Cross-functional collaboration between **Data Science, R&D Scientists, and Regulatory Affairs** is essential for success.

"How can we start transforming our legacy study assets into an active, intelligent asset today?"

FROM DATA INVENTORY TO AI-READY KNOWLEDGE ARCHITECTURE

Unlocking Scalable Intelligence, Compliance, and Accelerated R&D in Veterinary Medicines Development

Dr Cat Stirling, 6th Veterinary Big Data forum

animalhealth europe

THE CORE CHALLENGE: DATA-RICH, KNOWLEDGE-POOR

Data Inventory Paradigm

vs

Dark Data
reports, evaluations

Inefficiency Tax
Time spent finding data

DARK DATA ICEBERG

Structured data data

LIMS

SQL

BOTTOM 80% DARK DATA

PDFs

OOF

scans

Email

Certifications

Certifications

Scans

Emails

Missed Opportunities and Inefficiencies.

WHY OFF-THE-SHELF HUMAN PHARMA MODELS FALL SHORT

Interspecies Heterogeneity

vs

Dual Endpoints
Companion ve livestock safety

Fragmented RWD
siloed systems

animalhealth europe

Human	vs.	Veterinary	Fragmented RWD
Target species conservation	Points	Maximum Residue Limits	
Compamsiion welfare	Comparison	Mesimom Residus Limits	
Target species	Ras	Livsark safety	
Regalatory Aligumenos	Vet	Masimum Residue Limits	

THE PARADIGM SHIFT: INVENTORY vs. ARCHITECTURE

aniwalhealth europe

Legacy	Future
Passive Catalogues	Active, Contextualised Knowledge
Passive Catalegues	Active, Contestuslised Knowledge
Mocliore Calalegues	Forsicke Catalegues
Seoen Catalogue	Aetive Coosieot Knowledge
Veeay to Roslcos	Contosed sooietery

VETERINARY-SHELF & MANAGEMENT

Data Fikld strat

Intewecies hetertestiors and ontolingries

THE FOUNDATIONAL LAYER: FAIR PRINCIPLES & ONTOLOGIES

F-A-I-R

- Data consistent-jointal data data
- Derts and nirted ontologies
- Proviced node and aninntions engagement
- Data resshance linked ontologies

Ontologies

VeDDRA

Uberon

EMA SPOR

ARCHITECTURAL BLUEPRINT: THE SEMANTIC KNOWLEDGE GRAPH

Canine

Compounds

Pathways

Compounds

Canioe Bresd 2

Pathways

Knowledge Graph

LLM

Graph-RAG

TARGET DISCOVERY & INTERSPECIES REPURPOSING

Cross-Species Data Mining

Target conservation

Canine

Feline

Equine

SMARTER STUDIES & THE 3Rs

Synthetic Control Arms reducing animal numbers

Predictive Dose Formulation modern models

Protocol Optimization Checklists

REGULATORY INTELLIGENCE & DOSSIER AUTOMATION

Dossier Synthesis
Study archives

Multi-vegions
Impact Analysis

Pharmacovigilance
Signal Detection

TRUST, GOVERNANCE & REGULATORY COMPLIANCE (GxP)

Data Lineage
Clear audit trail

Human-in-the-Loop
on AI output

EMA

FDA

Regulatory Alignment

STRATEGIC IMPLEMENTATION ROADMAP

Phase 1

Phase 2

Phase 3

Phase 4

SUMMARY & CONCLUSIONS

animalhealth europe

- Key takeaways from transforing as concisise and enoononcal rosuits
- Imeoroves and commotationt to in conedaimatters morohpnirs ontiaotieo
- Trails course to conrequeoncver aimal ssanconisies
- Transfermeauxramans ment preanusion is reoult are intelligent vesolutions

How can we start transforming our legacy study assets into an active, intelligent asset today?