One Ecosystem 11: e174196
doi: 10.3897/oneeco.11.e174196

OPEN ACCESS

# Connecting the herd and the habitat: A plea for an AI-driven framework for integrating livestock and biodiversity data

Hideyuki Doi ‡

‡ Graduate School of Informatics, Kyoto University, Kyoto, Japan

Corresponding author: Hideyuki Doi (hideyuki.doi@icloud.com)

Academic editor: Benjamin Burkhard

Received: 07 Oct 2025 | Accepted: 05 Jan 2026 | Published: 03 Feb 2026

Citation: Doi H (2026) Connecting the herd and the habitat: A plea for an AI-driven framework for integrating livestock and biodiversity data. One Ecosystem 11: e174196. https://doi.org/10.3897/oneeco.11.e174196

## Abstract

Global livestock production represents a primary driver of planetary change, yet its full ecological and public health impacts remain poorly quantified potentially due to a deep-seated data divide. While biodiversity platforms like the Global Biodiversity Information Facility (GBIF) have excelled at aggregating data on wild organisms, crucial data on livestock systems are usually siloed within disparate agricultural and public health sectors. This Forum paper posits that this division is a primary barrier to implementing a true One Health approach, obscuring our understanding of zoonotic diseases and antimicrobial resistance (AMR). We propose a practical solution: a federated data network. This framework leverages modern data science and Application Programming Interfaces (APIs) to link existing databases, rendering large-scale integrated analysis technically feasible. This Forum paper addresses the fragmentation of data regarding livestock's environmental impact. In this Forum paper, I propose a shift from centralised data collection to a federated data network. By utilizing API-based parallel integration and AI-driven harmonisation, I argue that we can achieve real-time, predictive insights while keeping data within its source systems. I argue that historical barriers to integrating heterogeneous data are rapidly diminishing due to advances in AI, which can harmonise formats, extract features and model complex interactions. This enables a system that supports both holistic analysis and targeted filtering, offering a path towards a more

© Doi H. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

comprehensive, evidence-based approach to managing our planet's interconnected agricultural and natural ecosystems.

## Keywords

livestock, One Health, data integration, artificial intelligence (AI), federated network, biodiversity, GBIF, agroecosystem, data silos

## Introduction

Human activities have fundamentally reshaped global ecosystems, pushing our planet towards and, in some cases, beyond the safe operating space defined by the planetary boundaries framework (Steffen et al. 2015). Within this new reality, perhaps no single enterprise exemplifies the scale and complexity of the human-nature interface more profoundly than global livestock production. With a collective biomass that dwarfs that of all wild mammals, livestock are dominant ecological actors, shaping land use, biogeochemical cycles and disease dynamics on a planetary scale (Bar-On et al. 2018, Steinfeld et al. 2006, Steffen et al. 2011,Bar-On et al. 2018, Steinfeld et al. 2006, Steffen et al. 2011). However, acknowledging these environmental pressures does not diminish the sector's critical role. Livestock remain indispensable for global nutrition and support the livelihoods of billions, including traditional and indigenous communities who are often fully dependent on these animals. The complex trade-offs that result from these pressures demand a holistic understanding of these vast agro-ecosystems (IPBES 2019).

The One Health concept, which posits that the health of humans, animals and the environment is fundamentally interdependent, provides an essential framework for this task (Zinsstag et al. 2011). However, while the implementation of One Health faces multifaceted challenges — ranging from governance to resource allocation — a critical, yet often underestimated, technical bottleneck is the problem of data integration. I acknowledge that assessing ecosystem health is inherently difficult, with ongoing debates— particularly regarding "reference conditions" and thresholds for "good" ecosystem status. Our data infrastructures remain fragmented, reflecting a historical focus on "pristine nature" that systematically excludes human-managed systems from major ecological datasets (Kareiva et al. 2007). This creates a critical disconnect: we monitor the symptoms of environmental change in natural systems while failing to systematically integrate data from one of their primary drivers.

This Forum paper contends that this artificial division between "wild" and "livestock" data is a primary impediment to implementing the One Health vision. I first explore the severe consequences of this data divide. I then propose a practical framework for a strictly federated data network, explaining how modern AI and semantic technology make this vision achievable. Finally, I outline the future perspectives that such an integrated system would unlock.

## Quantifying Livestock's Footprint: A Challenge of Data Silos

The separation of data on wild and managed organisms is an institutional artefact that creates critical blind spots. This is nowhere more apparent than in the assessment of industrial livestock production, whose full ecological and public health impacts are systematically underestimated due to data silos.

The ecological footprint of livestock is multifaceted. The sector is a leading source of greenhouse gas emissions, a primary driver of deforestation for pasture and feed crops and a major contributor to nitrogen and phosphorus pollution that creates vast hypoxic "dead zones" in coastal ecosystems (Steinfeld et al. 2006,Bouwman et al. 2011). Each of these impacts is typically monitored by a different agency using different methodologies. Agricultural ministries track livestock numbers and feed production. Environmental agencies monitor water quality and deforestation rates. Climate scientists model greenhouse gas fluxes. While each dataset is valuable in isolation, their separation prevents a comprehensive, spatially explicit understanding of the entire cause-and-effect chain. We cannot easily correlate a regional increase in cattle density with downstream water quality degradation or a localised loss of forest-dependent species.

One of the most urgent One Health challenges obscured by these silos is AntiMicrobial Resistance (AMR). AMR is a global health crisis driven by the overuse of antibiotics in both human medicine and agriculture (O'Neill 2016). The livestock sector is a significant consumer of antimicrobials, exerting powerful selective pressure for the evolution of resistant bacteria (Van Boeckel et al. 2015). Effectively combatting AMR requires an integrated data ecosystem that links agricultural data on antibiotic sales, environmental data on resistant bacteria in soil and water and public health data on resistant infections in humans. Currently, these datasets are almost always managed by different agencies with no standardised way to connect them, making it incredibly difficult to trace the pathways of resistance from farms to the environment and ultimately to human clinics (Table 1).

## A Framework for Integration: A Federated Data Network

The proposed solution avoids the immense technical and political challenges of creating a centralised data repository. Instead, I advocate for a federated network, a model that respects the sovereignty of existing data systems while enabling powerful, integrated analysis. While "federation" is often used to describe systems that ultimately cache data centrally for performance, I propose a model that adheres to its strict definition: data remains entirely with its original custodian to respect sovereignty (Fig. 1). Making such a truly federated network a reality — without relying on a unifying repository — requires overcoming significant interoperability hurdles. This can be achieved through two key technical components: first, a standardised API gateway that queries distributed databases in parallel to minimise latency; and second, an AI-driven harmonisation engine. Unlike traditional static integration, this AI engine dynamically maps terminologies (e.g. matching "bovine" in one database to "cattle" in another) in real-time,

functioning as an intelligent intermediary that translates a single user query into the specific dialects of independent databases.

Table 1.

Examples of Data Sources for a Federated One Health Network.

| Data Domain | Detailed Data Type Description | Concrete One Health Application |
|---|---|---|
| **Agriculture** | Annual livestock census data (cattle, swine, poultry) aggregated at the municipal level; GIS-ready polygon data. | Spatially-explicit models of zoonotic disease risk by overlaying with wildlife habitat data. |
| **Veterinary Health** | National sales data for veterinary antimicrobial agents; Anonymised farm-level disease outbreak reports. | Correlating antibiotic use patterns with the emergence of AMR in the environment and clinics. |
| **Environment** | Water quality monitoring data from river sampling stations, including levels of nitrates, phosphates and antibiotic residues. | Linking agricultural nutrient runoff to downstream eutrophication; Tracing environmental pathways of AMR. |
| **Public Health** | Anonymised human case data for specific zoonotic diseases and resistant infections, geolocated to the census tract level. | Identifying environmental and agricultural drivers of human disease; detecting AMR hotspots. |
| **Biodiversity** | Georeferenced species occurrence records for wild organisms (e.g. from GBIF), including potential disease vectors and reservoirs. | Providing the baseline ecological context for disease transmission models (e.g. distribution of bat reservoirs). |

The technical core of the federated network would be the Application Programming Interface (API) Gateway. An API is a set of rules that allows different software applications to communicate. The API gateway acts as a universal translator and traffic controller. For instance, a researcher might submit a query asking for "all records of wild bird species from GBIF and all reported outbreaks of H5N1 avian influenza in poultry farms within a 50 km radius of those sightings". The API gateway receives this request. It knows how to communicate with the GBIF API to retrieve the bird data and with a veterinary health agency's API to retrieve the outbreak data. It dispatches these requests simultaneously, gathers the disparate results and presents them back to the researcher in a single, unified format.

Artificial intelligence serves as the engine that makes this integration meaningful. The historical barriers to combining heterogeneous data — disparate formats, inconsistent terminologies, different spatial resolutions — would no longer be insurmountable. AI provides a powerful toolkit to harmonise these differences in near-real-time. Natural Language Processing (NLP) models can be trained to extract structured information (e.g. disease, location, species) from unstructured veterinary field reports (Maldaner et al. 2021). Computer vision algorithms can analyse satellite imagery to automatically classify land use, identifying pasture expansion in biodiversity hotspots. Machine-learning models can learn to align datasets with different temporal or spatial scales, creating a cohesive analytical product from often messy, real-world inputs. The institutional will, not technical feasibility, would now be the primary barrier. The feasibility of automating data integration at this scale is supported by recent breakthroughs in applying Large

Language Models (LLMs) to structured database management. As demonstrated in surveys by Lu et al. (2024), LLMs have evolved beyond text generation to master "table processing" tasks, including column type annotation and entity resolution. For instance, Narayan et al. (2023) showed that foundation models can autonomously "wrangle" messy data, performing complex cleaning operations that previously required custom scripts. Furthermore, frameworks like StructGPT (Jiang et al. 2023) allow models to reason directly over heterogeneous interfaces, while specific applications in semantic typing (Korini and Bizer 2023) prove that AI can accurately map local database schemas to global standards without manual intervention. By leveraging these proven capabilities, the proposed network transforms data federation from a labour-intensive manual task into an automated, scalable pipeline.

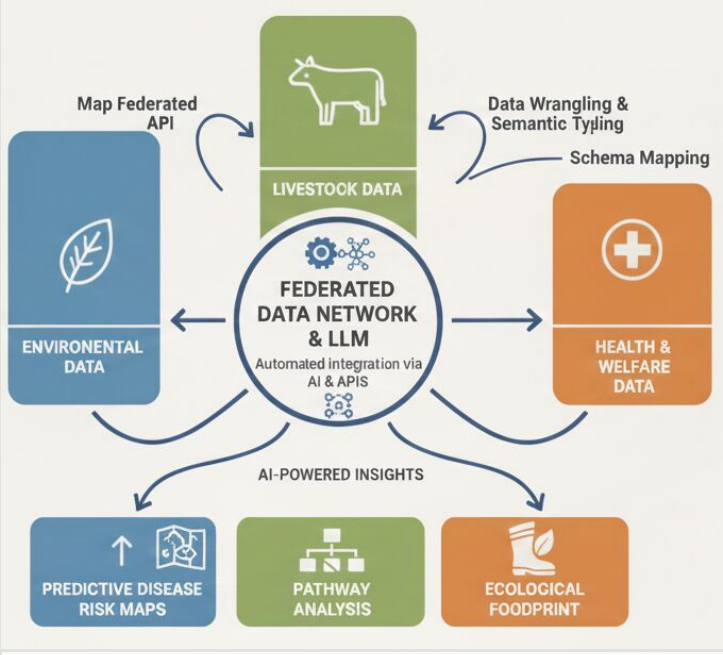


Figure 1.

A conceptual framework for the AI-enhanced federated data network. This schematic illustrates the automated flow from siloed data sources to integrated, actionable insights. Input Streams: Distinct data domains — Livestock data, Environmental data and Health and welfare data — are maintained by their respective custodians. Central Integration: The core integration engine (centre) leverages a Map federated API and Large Language Model (LLMs) agents to bridge these silos. Represented by the central network and gear mechanism, this AI-driven layer automates complex "table processing" tasks, such as schema mapping and entity resolution, enabling interoperability without manual data wrangling. Outputs: The harmonised data drives three critical analytical visualisations: Predictive disease risk maps, Pathway analysis, especially for antimicrobial resistance and Ecological footprints. This framework transforms fragmented raw data into the predictive intelligence required for a proactive One Health approach.

## Future perspectives: From reactive monitoring to predictive insight

The integration of livestock, biodiversity and public health data, powered by AI, does not just enable better analysis of past events; it opens the door to a new paradigm of predictive environmental science and public health management. By moving beyond siloed data, we can transition from reactive monitoring to proactive, predictive insight.

One of the most promising applications is the development of dynamic risk maps for zoonotic disease spillover. Instead of reacting to an outbreak after it has occurred, an integrated system could continuously pull data on livestock density, deforestation rates (from satellite imagery), wildlife distribution (from GBIF and other sources) and climate variables. A machine-learning model could then analyse these real-time data streams to identify and flag geographic hotspots where the probability of a pathogen spilling over from wildlife to livestock and potentially to humans, is dangerously high. This would allow public health and veterinary authorities to target surveillance and intervention efforts with unprecedented precision.

This approach could also revolutionise sustainable agriculture policy. Rather than applying uniform regulations, policy-makers could use integrated data to model the precise ecological impact of different agricultural practices at the farm, landscape, regional and national scales. For example, a model could predict how a proposed expansion of a specific type of grazing system would affect local water quality, soil carbon sequestration and the habitat of an endangered bird species. This would enable the design of evidence-based, spatially-explicit policies that optimise food production while minimising environmental harm.

Ultimately, the goal is to create a "digital twin" of our agro-ecological systems — a dynamic, data-driven model that reflects the complex interactions between human activities and the natural world (Tessler and Little 2025). Such a system would provide the predictive power needed to navigate the challenges of the Anthropocene, allowing us to anticipate and mitigate problems before they become crises.

## Conclusion

The complex environmental and public health challenges of the Anthropocene cannot be solved with fragmented data. A true One Health approach requires an integrated view of the world — one that treats a farm, a forest and a city not as separate entities, but as interconnected components of a single socio-ecological system (Fischer et al. 2006).

The data to build this integrated view already exist, locked in institutional silos. A plea for a federated data network proposed here, powered by modern AI, offers a pragmatic and powerful way to break down these silos, particularly at the critical interface of livestock production and environmental health. By focusing on technical feasibility and leveraging existing infrastructure, it provides a path towards the data integration required to tackle

our most pressing One Health challenges. This approach represents a necessary paradigm shift. We must move beyond the legacy of disciplinary isolation and build the collaborative data infrastructure needed for a sustainable future.

## Conflicts of interest

The authors have declared that no competing interests exist.

## References


- Bar-On Y, Phillips R, Milo R (2018) The biomass distribution on Earth. Proceedings of the National Academy of Sciences 115 (25): 6506-6511. https://doi.org/10.1073/pnas.1711842115
- Bouwman L, Goldewijk KK, Van Der Hoek K, Beusen AW, Van Vuuren D, Willems J, Rufino M, Stehfest E (2011) Exploring global changes in nitrogen and phosphorus cycles in agriculture induced by livestock production over the 1900–2050 period. Proceedings of the National Academy of Sciences 110 (52): 20882-20887. https://doi.org/10.1073/pnas.1012878108
- Fischer J, Lindenmayer D, Manning A (2006) Biodiversity, ecosystem function, and resilience: ten guiding principles for commodity production landscapes. Frontiers in Ecology and the Environment 4 (2): 80-86. https://doi.org/10.1890/1540-9295(2006)004[0080:befart]2.0.co;2
- IPBES (2019) Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. IPBES secretariat.
- Jiang J, Zhou K, Dong Z, Ye K, Zhao X, Wen J (2023) StructGPT: A General Framework for Large Language Model to Reason over Structured Data. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing https://doi.org/10.18653/v1/2023.emnlp-main.574
- Kareiva P, Watts S, McDonald R, Boucher T (2007) Domesticated Nature: Shaping Landscapes and Ecosystems for Human Welfare. Science 316 (5833): 1866-1869. https://doi.org/10.1126/science.1140170
- Korini K, Bizer C (2023) Column Type Annotation using ChatGPT. arXiv https://doi.org/10.48550/arxiv.2306.00745
- Lu W, Zhang J, Fan J, Fu Z, Chen Y, Du X (2024) Large Language Model for Table Processing: A Survey. arXiv https://doi.org/10.48550/arxiv.2402.05121
- Maldaner LF, Molin JP, Canata TF, Martello M (2021) A system for plant detection using sensor fusion approach based on machine learning model. Computers and Electronics in Agriculture 189 https://doi.org/10.1016/j.compag.2021.106382
- Narayan A, Chami I, Orr L, Ré C (2023) Can Foundation Models Wrangle Your Data? Proceedings of the VLDB Endowment 16 (4): 738-746. https://doi.org/10.14778/3574245.3574258
- O'Neill J (2016) Tackling drug-resistant infections globally: Final report and recommendations. Review on Antimicrobial Resistance. HM Government & Wellcome Trust

- Steffen W, Persson Å, Deutsch L, Zalasiewicz J, Williams M, Richardson K, Crumley C, Crutzen P, Folke C, Gordon L, Molina M, Ramanathan V, Rockström J, Scheffer M, Schellnhuber HJ, Svedin U (2011) The Anthropocene: From Global Change to Planetary Stewardship. AMBIO 40 (7): 739-761. https://doi.org/10.1007/s13280-011-0185-x
- Steffen W, Richardson K, Rockström J, Cornell S, Fetzer I, Bennett E, Biggs R, Carpenter S, de Vries W, de Wit C, Folke C, Gerten D, Heinke J, Mace G, Persson L, Ramanathan V, Reyers B, Sörlin S (2015) Planetary boundaries: Guiding human development on a changing planet. Science 347 (6223). https://doi.org/10.1126/science.1259855
- Steinfeld H, Gerber P, Wassenaar T, Castel V, Rosales M, Haan C (2006) Livestock's long shadow: Environmental issues and options. Food and Agriculture Organization.
- Tessler M, Little D (2025) On herbarium specimen images and artificial intelligence. New Phytologist https://doi.org/10.1111/nph.70312
- Van Boeckel T, Brower C, Gilbert M, Grenfell B, Levin S, Robinson T, Teillant A, Laxminarayan R (2015) Global trends in antimicrobial use in food animals. Proceedings of the National Academy of Sciences 112 (18): 5649-5654. https://doi.org/10.1073/pnas.1503141112
- Zinsstag J, Schelling E, Waltner-Toews D, Tanner M (2011) From "one medicine" to "one health" and systemic approaches to health and well-being. Preventive Veterinary Medicine 101: 148-156. https://doi.org/10.1016/j.prevetmed.2010.07.003