Journal of Integrative Science and Societal Impact
July-September 2026 2(2)
https://jissi.org

Original Article

# FROM STATISTICIAN-DEPENDENT PRACTICE TO DATA-ANALYTIC AGENCY: AN ACTIVITY-THEORETICAL CASE FOR TRAINING VETERINARY AND HEALTH PROFESSIONALS AS DATA ANALYSTS IN THE AGE OF AI

## A SYSTEMIC-STRUCTURAL ACTIVITY THEORY ARGUMENT FOR INTEGRATED DATA WORK IN VETERINARY MEDICINE, ANIMAL HUSBANDRY, AND HEALTH SCIENCES

**Federico de la Colina Flores** [1*], **Paul Alexis de la Colina García** [2], **Tzitzi de la Colina García** [3], **Heriberto Rodríguez Frausto** [4]
[1] Universidad Autónoma de Zacatecas, México
[2] Colegio Edison A.C. Guadalupe, Zacatecas, México
[3] Colegio Edison A.C. Guadalupe, Zacatecas, México
[4] Universidad Autónoma de Zacatecas, México

## ABSTRACT

Veterinary medicine, animal husbandry, and health sciences operate in data-saturated ecologies shaped by electronic records, sensor systems, laboratory platforms, and increasingly automated analytic infrastructures. In this context, the dominant educational dilemma is not whether professionals should be statistically literate, but whether they should remain dependent on specialist analysts for the core analytic actions that define contemporary evidence-based practice, production optimization, and welfare governance. This article argues that training veterinarians, zootechnists, and health professionals as data analysts is superior to training them primarily as statisticians or as researchers who routinely outsource analytic work. The argument is grounded in cultural-historical activity theory and Bedny's systemic-structural activity theory, and it is updated for a technical environment in which machine learning, large language models, and agentic workflows reduce the friction of coding and documentation while increasing the need for domain-grounded judgement. R, RStudio, and Quarto are presented as an exemplary toolchain that unifies analysis, explanation, and publication into a single reproducible activity. The conclusion proposes a competence-oriented curriculum in which communicative, transformative, and evaluative capabilities co-develop through authentic analytic work tied to welfare and community responsibility.

**Keywords:** Activity Theory, Systemic-Structural Activity Theory, Data Literacy, Veterinary Informatics, R, RStudio, Quarto, Reproducible Research, Machine Learning, Large Language Models

## INTRODUCTION

Across veterinary practice, animal production systems, and health research, decisions increasingly arise from complex data flows rather than isolated observations. Precision livestock farming and digital monitoring systems generate continuous measurements that demand model-based interpretation, and the long-term success of these systems depends on workforce capacity in programming and data science as much as on traditional animal science knowledge Menendez et al. (2022). Veterinary medicine has similarly faced an informatics imperative: effective care and research require not only data capture but interoperable

*Corresponding Author:
**Email address:** Federico de la Colina Flores (federcol@hotmail.fr)

**Received:** 11 June 2026; **Accepted:** 13 July 2026; **Published** 06 August 2026
**DOI:** 10.29121/JISSI.v2.i2.2026.35
**Page Number:** 23-28
**Journal Title:** Journal of Integrative Science and Societal Impact
**Journal Abbreviation:** J. Integr. Sci. Soc. Impact
**Online ISSN:** 3108-2165, **Print ISSN:** 3108-1959
**Publisher:** Granthaalayah Publications and Printers, India
**Conflict of Interests:** The authors declare that they have no competing interests.
**Funding:** This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.
**Authors' Contributions:** Each author made an equal contribution to the conception and design of the study. All authors have reviewed and approved the final version of the manuscript for publication.
**Transparency:** The authors affirm that this manuscript presents an honest, accurate, and transparent account of the study. All essential aspects have been included, and any deviations from the original study plan have been clearly explained. The writing process strictly adhered to established ethical standards.

information models, analytic capacity, and collaboration across disciplines Johnson et al. (2011). These realities make analytic work an ordinary component of professional action rather than an occasional technical service.

Yet many curricula still treat data analysis as an accessory. A common compromise is to strengthen statistics courses while leaving authentic analytic work to later specialist consultation. Another compromise is to train researchers to design studies but to outsource data handling, programming, and reproducible reporting. These arrangements can produce competent consumers of results, but they often leave the professional unable to carry the object of practice through the full chain of analytic transformation, from raw records to warranted claims. In welfare-relevant domains, such dependence is costly. It delays feedback, increases miscommunication, obscures accountability, and weakens the professional's capacity to justify decisions to clients, producers, regulators, and communities.

The thesis advanced here is that training veterinarians, zootechnists, and health professionals as data analysts is superior because it reorganizes the activity system of professional formation. It makes analytic action an internal capability of the collective subject rather than an external service. It also aligns training with the historical trajectory of data analysis as a field that exceeds classical statistics in scope, integrating computational practices, visualization, domain reasoning, and iterative model use Donoho (2017), Tukey (1962). This superiority is not a rejection of statistics. It is a claim about division of labor and the location of analytic agency within professional activity.

The contemporary environment strengthens this case. Machine learning and large language models increasingly automate aspects of coding and documentation, and systematic reviews show that such models can enhance code completion and developer productivity while changing what it means to "write" software Husein et al. (2025), Jiang et al. (2024), Weber et al. (2024). The easier code becomes to draft, the more important it becomes for domain professionals to know what to ask, what to check, and how to integrate outputs into ethical and scientific decision-making. Analytic agency therefore becomes a central educational objective.

## CONCEPTUAL FRAMEWORK: ACTIVITY THEORY, MODIFIED ACTIVITY SYSTEMS, AND SSAT

Cultural-historical activity theory treats human practice as object-oriented, mediated, and collective. The unit of analysis is not an isolated individual act but a historically formed system in which subjects pursue objects through tools within communities and divisions of labor Engeström (1987), Leontiev (1978). On this view, education should not merely transmit information but reorganize the learner's participation in object-directed activity so that the learner becomes capable of acting within, and transforming, real practices.

Bedny's systemic-structural activity theory adds a fine-grained analytic language for how activity is organized and regulated. SSAT conceptualizes activity as a self-regulated system integrating cognitive, behavioral, and motivational components oriented toward conscious goals, and it provides a principled account of the relations among activity, actions, and operations Bedny and Karwowski (2006). This matters for education because "learning" becomes the development and reorganization of goal-directed actions and their operationalization under changing conditions, not merely the acquisition of declarative knowledge.

This article also assumes a modified activity system concept developed for your educational model, where the classical "rules" mediator is replaced by a persuasion tool. In professional formation, norms rarely operate as inert constraints. They are enacted through communicative artifacts, justificatory genres, ethical commitments, and institutional rhetorics that mobilize pathos, logos, and ethos. When analytic work is brought inside the collective subject, persuasion tools become essential because analytic claims must be translated into trust, consent, compliance, and coordinated action without coercion. This is especially salient in veterinary and husbandry contexts where decisions intersect with owner values, producer constraints, animal welfare, and public health.

## CLARIFYING THE ALTERNATIVES: STATISTICIAN TRAINING, CONSULTANT DEPENDENCE, AND DATA-ANALYTIC FORMATION

### THE ARGUMENT REQUIRES A CLEAR DISTINCTION AMONG THREE EDUCATIONAL ENDPOINTS.

**A statistician-oriented** endpoint emphasizes formal inference, probability models, and methodological rigor. This training is invaluable, but in domain programs it often remains detached from full data workflows, including cleaning, provenance, visualization, version control, and reproducible reporting. It can produce professionals who understand tests and models yet cannot enact analytic work reliably under real constraints.

**A consultant-dependent** endpoint trains professionals to design studies and interpret results while routinely delegating data handling, programming, and modeling implementation to specialized analysts. This can work in high-resource environments, but it introduces systematic vulnerabilities: translation loss between domain and analytic languages, delays in iterative inquiry, and diffusion of responsibility when results are contested.

**A data-analytic** endpoint trains professionals to perform end-to-end analytic actions as part of their normal object work. In this endpoint, statistics remains present as a warranting discipline, but it is embedded within broader data practices. The professional

learns to produce, critique, and revise models; to document decisions; to communicate uncertainty; and to deploy analytic artifacts as mediators of collective decision-making. This endpoint aims at agency, not merely literacy.

The superiority claim defended here is that the data-analytic endpoint better fits the historical development of knowledge production and the contemporary tool environment, and it better supports welfare-oriented practice in veterinary and health domains.

## WHY DATA-ANALYTIC TRAINING IS SUPERIOR IN WELFARE-ORIENTED DOMAINS

### ANALYTIC AGENCY PRESERVES THE INTEGRITY OF THE OBJECT OF ACTIVITY

In activity theory, the object is not simply a target. It is the meaningful problem-space that organizes motives, goals, and evaluation. In veterinary and husbandry practice, the object often includes an inseparable bundle: animal health, welfare, owner livelihood, environmental constraints, and institutional accountability. When analytic work is outsourced, the object risks fragmentation. Analysts may optimize for statistical elegance, while clinicians optimize for timeliness, and producers optimize for cost. A data-analytic professional can keep the object coherent across the analytic cycle because the same subject who understands the lived contradictions also performs the transformations that turn records into reasons.

This coherence becomes increasingly important as data sources multiply. Precision livestock systems, for example, require iterative modeling aligned with management goals and technology capabilities, and the literature explicitly frames workforce training in programming and data science as imperative for long-term success Menendez et al. (2022). Where the object is dynamic, analytic dependence becomes a bottleneck. Data-analytic agency reduces this bottleneck by enabling rapid hypothesis revision and context-aware model adjustment.

### ANALYTIC AGENCY REDUCES TRANSLATION LOSS AND SUPPORTS ACCOUNTABLE JUSTIFICATION

Consultant dependence introduces a structural translation problem. The professional must express domain questions as analytic specifications and then translate outputs back into domain decisions. Each translation step is an opportunity for misunderstanding, selective emphasis, and loss of nuance. This is not merely an efficiency issue. It affects accountability. In contested decisions, such as euthanasia, antimicrobial stewardship, or herd-level interventions, justification must be traceable and communicable. Evidence-based veterinary medicine pedagogy recognizes that clinical action depends on integrating research evidence with professional judgement and client values, and it calls for explicit training in accessing and applying knowledge through practice-centered exercises Fajt et al. (2009). Data-analytic training strengthens this integration because it places the professional in direct contact with the evidentiary transformations that generate the "evidence" they cite.

In SSAT terms, outsourcing pushes crucial goal-directed actions outside the learner's regulated activity, leaving the professional with partial control of the functional system that produces claims. A data-analytic endpoint internalizes these actions. It enables the professional to defend warrants, recognize limitations, and revise methods without waiting for external cycles.

### ANALYTIC AGENCY STRENGTHENS THE COLLECTIVE SUBJECT AND ITS DIVISION OF LABOR

Activity systems are collective. The question is not whether there is division of labor, but how it is organized. Data-analytic formation does not eliminate specialists. It changes the social topology of collaboration. When domain professionals possess analytic competence, collaboration with statisticians and data scientists shifts from dependency to partnership. Discussions become about model assumptions, decision consequences, and system constraints, rather than about basic feasibility or elementary coding. This produces a more inclusive collective intelligence, where specialized expertise amplifies the group rather than substituting for missing capacities.

Veterinary informatics scholarship frames data and information management as essential for interdisciplinary collaboration and institutional effectiveness Johnson et al. (2011). Data-analytic formation operationalizes this imperative at the level of training: learners can participate meaningfully in informatics and analytics communities rather than remaining peripheral consumers of outputs.

### ANALYTIC AGENCY SUPPORTS REPRODUCIBILITY AND CUMULATIVE KNOWLEDGE

Scientific work depends on cumulative reliability. Reproducible research principles emphasize that analyses must be transparent, executable, and inspectable, because hidden steps and undocumented choices undermine trust and hinder reuse Peng (2011), Sandve et al. (2013). Literate programming extends this logic by integrating narrative and computation into a single artifact so that explanations and results remain coupled Knuth (1984). In welfare-relevant fields, reproducibility is not a luxury. It is a condition for responsible policy, clinical governance, and research translation.

Data-analytic formation makes reproducibility part of the ordinary action repertoire of professionals. It trains them to produce analytic artifacts that can be audited, shared, and improved. This is also a development claim: students become capable of contributing to institutional memory and collective learning rather than producing isolated, non-reusable reports.

## THE ENABLING ENVIRONMENT: AI ASSISTANCE AND THE DECLINING MARGINAL COST OF CODING

The practical feasibility of data-analytic training has changed. Historically, a strong argument for analytic outsourcing was the high cost of programming skill acquisition. Today, machine learning systems and large language models reduce that cost by providing interactive code completion, translation from natural language to code, debugging suggestions, and scaffolded documentation. Systematic reviews report that large language models can significantly enhance code completion across languages and contexts, improving productivity by predicting relevant code based on partial input Husein et al. (2025). Controlled studies also report substantial productivity gains in programming tasks when developers use LLM-based assistance compared with traditional workflows Weber et al. (2024). Surveys synthesize a fast-growing body of work on LLMs for code generation and highlight both capabilities and risks, including correctness and evaluation challenges Jiang et al. (2024).

These systems do not eliminate the need for methodological competence. They shift the skill center of gravity. The scarce skill becomes the ability to formulate good questions, specify constraints, detect errors, evaluate plausibility, and connect results to domain objects and values. In other words, AI increases the returns to domain-grounded analytic agency. A professional who cannot judge outputs is vulnerable to automation bias. A professional who can judge outputs becomes dramatically more productive.

Agentic workflows add another step. LLM-based agents are increasingly designed to plan, call tools, and iterate toward solutions, extending LLM utility beyond text generation into structured problem solving and software tasks Liu et al. (2024). In educational terms, this suggests a new didactic opportunity: students can learn analytic activity by supervising semi-automated toolchains, focusing attention on warrants, assumptions, and consequences rather than on syntactic struggle.

## R, RSTUDIO, AND QUARTO AS AN INTEGRATED TOOLCHAIN FOR ANALYTIC FORMATION

The case for data-analytic training becomes concrete when the curriculum is anchored to a coherent tool ecology.

R remains a foundational environment for statistical computing and data analysis, widely used across biomedical and agricultural sciences R Core Team. (2025). The tidyverse ecosystem organizes common data tasks into a consistent grammar that supports learning, transparency, and communicability in analytic work Wickham et al. (2019), Wickham and Grolemund (2023). RStudio offers an integrated development environment that supports scripting, visualization, project organization, and reproducible workflows, and it is explicitly designed to improve productivity in R-based analysis Allaire (2011), Posit Team. (2025). Quarto extends literate programming into a multi-format publishing system that can render the same analytic manuscript into HTML, PDF, and Word while preserving executable code and citations Knuth (1984), Quarto Development Team. (2025).

For professional formation, the value of this toolchain is structural. It supports three competency dimensions that align with a theoretical, practical, poitetic educational model.

Communicative competence develops because Quarto requires the learner to explain. Code, results, and narrative are co-produced, making communication part of production rather than a post hoc add-on. Transformative competence develops because learners manipulate real data, build models, and generate artifacts that change practice, whether through dashboards, reports, or decision aids. Evaluative competence develops because the workflow foregrounds validation, uncertainty, sensitivity analysis, and the comparison of competing explanations.

This toolchain also supports the ethical dimension of professional responsibility. In veterinary and animal husbandry settings, analytic outputs must be defensible to multiple stakeholders. Quarto-based documents preserve an audit trail. They show what was done, with what data, and why particular transformations were chosen. This strengthens welfare governance because it supports reason-giving rather than authority claims.

## CURRICULUM IMPLICATIONS IN ACTIVITY-THEORETICAL TERMS

A data-analytic curriculum should be designed as progressive participation in object-oriented analytic activity. The learner begins with guided actions that are explicit and slow, and over time these actions become operationalized, conditionalized, and transferable across contexts, consistent with the activity–action–operation hierarchy emphasized in SSAT Bedny and Karwowski (2006). In early phases, learners work with small, well-understood datasets and produce short, reproducible reports. In later phases, they engage with noisy institutional data, sensor streams, and mixed-method evidence, integrating domain judgement with computational methods.

The central pedagogical unit is not the lecture on a statistical test. It is the analytic task embedded in a real object. A pharmacology case becomes an analytic problem about treatment response and uncertainty. A herd health rotation becomes an analytic problem about risk, surveillance, and intervention priorities. A wildlife health project becomes an analytic problem about

ecological constraints and ethics. Evidence-based veterinary medicine exercises already provide a template: practice-centered rounds build competence by repeatedly linking questions, evidence, and decisions Fajt et al. (2009). Data-analytic formation generalizes this template by making data work the medium through which questions become answerable and decisions become justifiable.

Within our modified activity system, persuasion tools also become curricular objects. Students must learn to translate analytic results into ethically sound persuasion that avoids manipulation, scientism, and coercion. Analytic competence therefore includes rhetorical competence: how to justify a model choice to a producer, how to explain uncertainty to a client, how to argue for welfare safeguards to an institution, and how to document decisions in ways that invite critique.

## ANTICIPATING OBJECTIONS AND SPECIFYING THE CONDITIONS OF SUPERIORITY
## TWO OBJECTIONS DESERVE CAREFUL TREATMENT.

**First,** it may be argued that data-analytic training dilutes statistical rigor. The correct response is to treat statistics as the warranting core of analytic judgement while refusing to reduce analytic activity to inference alone. The history of data analysis emphasizes exploration, visualization, and iterative sense-making as integral to reliable work, not as optional preliminaries Donoho (2017), Tukey (1962). Data-analytic formation can increase rigor by making assumptions explicit in code and by preserving decision trails. It also improves the professional's ability to know when a statistician is needed, and to collaborate effectively when advanced methods are required. Kasarapu (2026)

**Second,** it may be argued that specialist analysts are more efficient. This is true when analytic work is routine, well-specified, and separable from domain action. It is less true when the object is dynamic, when feedback cycles are rapid, and when values and constraints shift during inquiry. Veterinary and health practice often exhibits precisely these conditions. Under such conditions, consultant dependence becomes fragile. The superiority claim defended here is therefore conditional but robust: in data-rich, welfare-relevant domains where decisions are iterative and contested, embedding analytic agency within the professional collective subject produces better outcomes than outsourcing core analytic actions.

## CONCLUSION

Training veterinarians, zootechnists, and health professionals as data analysts is superior because it reorganizes professional formation around analytic agency. It preserves the integrity of the object of activity, reduces translation loss, strengthens accountable justification, and supports reproducible, cumulative knowledge. It also reconfigures collaboration from dependency to partnership, enabling collective intelligence in interdisciplinary environments.

The argument is strengthened by current technical conditions. Machine learning, large language models, and agentic workflows reduce the marginal cost of coding while increasing the value of domain-grounded judgement. R, RStudio, and Quarto provide a coherent tool ecology for turning these conditions into educational practice, unifying analysis, explanation, and dissemination into a single reproducible activity.

In activity-theoretical terms, the aim is developmental: to transform the learner's role in the activity system so that analytic actions become part of professional identity and operational repertoire. In welfare-oriented fields, this is not merely a technical upgrade. It is a democratic responsibility, because it enables professionals to offer reasons that communities can inspect, contest, and improve. Hazarika. (2025)

## ACKNOWLEDGMENTS

None.

## REFERENCES

Allaire, J. J. (2011). RStudio: Integrated Development Environment for R. User! 2011 Conference Abstracts.

Bedny, G., and Karwowski, W. (2006). A Systemic-Structural Theory of Activity: Applications to Human Performance and Work Design. CRC Press. https://doi.org/10.1201/9781420009743

Donoho, D. (2017). 50 Years of Data Science. Journal of Computational and Graphical Statistics, 26(4), 745–766. https://doi.org/10.1080/10618600.2017.1384734

Engeström, Y. (1987). Learning by Expanding: An Activity-Theoretical Approach to Developmental Research. Orienta-Konsultit.

Fajt, V. R., Brown, D., and Scott, M. M. (2009). Practicing the Skills of Evidence-Based Veterinary Medicine Through Case-Based Pharmacology Rounds. Journal of Veterinary Medical Education, 36(2), 186–195. https://doi.org/10.3138/jvme.36.2.186

Hazarika, I. (2025). Impact of Consumer Analytics and AI on Inventory Management in UAE with Reference to UN SDG 12: Sustainable Consumption and Production Patterns and SDG 2 Zero Hunger in UAE. In 2025 IEEE Integrated STEM Education Conference (ISEC) (1–8). IEEE. https://doi.org/10.1109/ISEC64801.2025.11460353

Jiang, J., et al. (2024). A Survey on Large Language Models for Code Generation. ACM Computing Surveys. Advance Online Publication.

Jiang, R. A., Aburajouh, H., and Catal, C. (2025). Large Language Models for Code Completion: A Systematic Literature Review. Computer Standards and Interfaces. Advance Online Publication. https://doi.org/10.1016/j.csi.2024.103917

Johnson, L. M., Ames, T. R., Jacko, J. A., and Watson, L. A. (2011). The Informatics Imperative in Veterinary Medicine: Collaboration Across Disciplines. Journal of Veterinary Medical Education, 38(1), 5–9. https://doi.org/10.3138/jvme.38.1.5

Kasarapu B. C. (2026). Generative AI-Enabled Micro-Frontend Framework for Scalable and Intelligent Enterprise Retail Applications. International Journal of Computer Information Systems and Industrial Management Applications, 18(5s), 297–309. https://doi.org/10.70917/ijcisim-2026-2708

Knuth, D. E. (1984). Literate Programming. The Computer Journal, 27(2), 97–111. https://doi.org/10.1093/comjnl/27.2.97

Leontiev, A. N. (1978). Activity, Consciousness, and Personality. Prentice-Hall.

Liu, J., Wang, K., Chen, Y., Peng, X., Chen, Z., Zhang, L., and Lou, Y. (2024). Large Language Model-Based Agents for Software Engineering: A Survey (Arxiv Preprint).

Menendez, H. M., Brennan, J. R., Gaillard, C., Ehlert, K., Quintana, J., Neethirajan, S., Remus, A., Jacobs, M., Teixeira, I. A. M. A., Turner, B. L., and Tedeschi, L. O. (2022). ASAS-NANP Symposium: Mathematical Modeling in Animal Nutrition: Opportunities and Challenges of Confined and Extensive Precision Livestock Production. Journal of Animal Science, 100(6), Article Skac160. https://doi.org/10.1093/jas/skac160

Peng, R. D. (2011). Reproducible Research in Computational Science. Science, 334(6060), 1226–1227. https://doi.org/10.1126/science.1213847

Posit Team. (2025). RStudio: Integrated Development Environment for R (Version 2025) [Computer Software]. Posit Software, PBC.

Quarto Development Team. (2025). Quarto: Scientific and Technical Publishing System [Computer Software].

R Core Team. (2025). R: A Language and Environment for Statistical Computing [Computer Software]. R Foundation for Statistical Computing.

Sandve, G. K., Nekrutenko, A., Taylor, J., and Hovig, E. (2013). Ten Simple Rules for Reproducible Computational Research. PLoS Computational Biology, 9(10), Article e1003285. https://doi.org/10.1371/journal.pcbi.1003285

Tukey, J. W. (1962). The Future of Data Analysis. The Annals of Mathematical Statistics, 33(1), 1–67. https://doi.org/10.1214/aoms/1177704711

Weber, T., Brandmaier, M., Schmidt, A., and Mayer, S. (2024). Significant Productivity Gains Through Programming with Large Language Models. In Proceedings of the 16th ACM SIGCHI Symposium on Engineering Interactive Computing Systems. Association for Computing Machinery. https://doi.org/10.1145/3661145

Wickham, H., and Grolemund, G. (2023). R for Data Science (2nd ed.). O'Reilly Media.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L. D., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T. L., Miller, E., Bache, S. M., Müller, K., Ooms, J., Robinson, D., Seidel, D. P., Spinu, V., ... Yutani, H. (2019). Welcome to the Tidyverse. Journal of Open Source Software, 4(43), Article 1686. https://doi.org/10.21105/joss.01686