Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Digital Athlete Sovereignty: Navigating the Commodification and Ethics of Youth Football Data in a Globalized Ecosystem

Mona Suliman

monaawadsuliman@gmail.com

University of science and technology omdarmam

Systematic Review

Keywords: Youth Football, Data Commodification, Sports Analytics, Blockchain, Digital Labor, Athlete Privacy, Talent Identification, Surveillance Capitalism, Algorithmic Bias, Digital Athlete Sovereignty

Posted Date: July 6th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10229564/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

The global youth football ecosystem is undergoing a rapid digital transformation in which player performance metrics, biometric signals, video footage, and developmental profiles are increasingly collected, standardized, and monetized through analytics platforms. This article examines the commodification of youth football data and its implications for athlete privacy, platform power, algorithmic bias, and value distribution. Using a structured secondary-source review combined with policy analysis, platform analysis, and illustrative case-based discussion, the study maps the data lifecycle from local collection through aggregation, analytics, and commercial use. It argues that the benefits of digital scouting and performance optimization are unevenly distributed: elite academies and global platforms capture the greatest technical and economic advantages, while young athletes and grassroots organizations often have limited bargaining power. The article develops the concept of Digital Athlete Sovereignty as a governance principle for youth sports data and proposes a decentralized data ownership framework built around transparent consent, interoperability, blockchain-supported auditability, smart-contract-based permissions, and fair revenue-sharing mechanisms. The proposed model is presented as a conceptual framework requiring future prototype development, field testing, and regulatory validation rather than as a fully implemented solution. The article contributes to debates on sports data governance, digital labor, youth protection, and ethical artificial intelligence in talent identification.

1. Introduction

The global youth football ecosystem has undergone a profound digital transformation, evolving from traditional grassroots development to a sophisticated data-driven industry. With over 55 million youth athletes participating in organized sports worldwide and the youth sports sector generating approximately $3 billion in annual revenue, the integration of digital technologies has fundamentally reshaped talent identification, development, and commercialization processes (Project Play, 2025).

This transformation is particularly evident in football, where data analytics platforms now process over 1.4 million data points per game, creating unprecedented opportunities for performance optimization while simultaneously raising critical questions about data ownership, privacy, and equitable value distribution (API4AI, 2025).

The commodification of youth football data represents a paradigm shift in sports economics. What began as simple performance metrics tracked by coaches has evolved into a complex digital commodity traded across global platforms. Hudl's acquisition of Wyscout exemplifies this trend, creating an integrated ecosystem that aggregates video and statistical data from over 600 leagues worldwide, serving as a comprehensive repository for football analysis (Hudl, 2025a). However, this rapid digitization has outpaced regulatory frameworks and ethical guidelines, leaving young athletes—often minors with limited bargaining power—vulnerable to exploitation.

This paper addresses a critical research gap: the intersection of technological advancement, economic opportunity, and ethical responsibility in youth football data commodification. While professional football's data analytics sector has received considerable academic attention, the grassroots-to-global pipeline remains understudied, particularly regarding how local academies can benefit economically without exploiting young players.

The research is especially timely given emerging regulatory developments, such as California Assembly Bill 734 and evolving GDPR compliance requirements for youth sports organizations (O'Connor's Law, 2025).

The study employs a multidisciplinary approach, combining technical analysis of digital platform infrastructure with economic modeling and ethical frameworks. By examining the complete data lifecycle—from pitch sensors to cloud storage to AI-driven analytics—this research provides actionable insights for stakeholders seeking to balance commercial interests with athlete protection.

The central problem addressed in this article is not whether youth football should use data-driven tools, but how these tools can be governed so that technological innovation does not deepen existing inequalities or transform minors into uncompensated data producers. The article therefore treats youth football data as simultaneously developmental information, personal information, and a potential economic asset.

1.1 Research Questions

1. How is youth football data commodified across digital scouting, analytics, and performance-monitoring platforms?
2. What ethical, legal, and cybersecurity risks arise from the collection and commercial use of minors' biometric and performance data?
3. How does the digital divide between elite academies and grassroots organizations shape talent identification and value distribution?
4. How can a decentralized data ownership model improve transparency, athlete control, and fair compensation in youth football data ecosystems?

1.2 Research Objectives

- To analyze the mechanisms through which youth football data is extracted, standardized, aggregated, and monetized.
- To examine the ethical, legal, economic, and cybersecurity risks associated with youth athlete data commodification.
- To evaluate how platform power and unequal access to analytics technologies affect grassroots and elite academies differently.

- To propose a conceptual governance framework for Digital Athlete Sovereignty based on consent, transparency, interoperability, and fair value distribution.

1.3 Original Contribution

- Theoretical contribution: the article connects data commodification, surveillance capitalism, digital labor, and youth sports governance in a single analytical framework.
- Practical contribution: it translates abstract data rights into concrete recommendations for academies, platforms, athletes, families, and policymakers.
- Policy contribution: it identifies gaps in current regulatory approaches to minors' sports data and proposes principles for youth sports data protection.
- Technical contribution: it presents a conceptual decentralized data ownership model using identity management, permissioned blockchain, smart contracts, and interoperable APIs.

2. Literature Review and Research Gap

2.1 Youth Football Analytics and Digital Talent Identification

Research on football analytics has traditionally concentrated on professional performance analysis, match prediction, player recruitment, and injury prevention. In youth football, however, similar technologies raise additional concerns because the players are often minors, the data may include sensitive biometric indicators, and long-term career effects may occur before athletes have meaningful bargaining power. Digital platforms expand the visibility of young players but also create new forms of dependency on proprietary databases and platform-specific metrics.

2.2 Data Commodification and Digital Labor

The literature on data commodification explains how personal, behavioral, and performance information can be transformed into a marketable asset. In youth football, young athletes generate the raw material of value through training, competition, physiological activity, and video-recorded performance. Yet the economic value of this data is usually captured by platforms, clubs, and commercial intermediaries rather than by the athletes and families who generate it.

2.3 Algorithmic Bias and Inequality

AI-assisted scouting tools may improve decision-making by reducing purely subjective evaluation, but they may also reproduce historical inequalities if the training data overrepresents players from well-funded academies and underrepresents grassroots contexts. Metrics that appear objective can conceal unequal access to coaching, nutrition, facilities, competition exposure, and digital tracking infrastructure.

2.4 Legal and Ethical Governance of Minors' Sports Data

Existing data protection laws provide general protections for personal data and, in some jurisdictions, special protections for minors. However, youth sports data introduces complex questions that are not always directly addressed by general privacy regimes: who owns performance data, how long it may be retained, whether it can be monetized, and whether athletes should share in commercial value generated from their own biometric and performance profiles.

2.5 Research Gap

The key gap is the absence of an integrated governance framework that connects platform economics, data rights, minor protection, grassroots development, and technical architecture. Existing discussions often treat sports analytics as a performance tool, privacy law as a compliance issue, or blockchain as a technical solution. This article brings these debates together and evaluates whether Digital Athlete Sovereignty can serve as a practical normative framework for youth football data governance.

3. Methodology

3.1 Research Design

This study adopts a qualitative, non-empirical research design based on a structured secondary-source review, policy analysis, platform analysis, and illustrative case-based examination. It does not involve interviews, surveys, experiments, human-subject recruitment, or access to private athlete data. The methodological choice is appropriate because the article aims to develop a conceptual and governance framework rather than to test a causal hypothesis or evaluate a deployed technical system.

To avoid overstating the evidence base, the article is framed as a structured narrative and conceptual review rather than a formal PRISMA systematic review. A formal PRISMA review would require a complete database search log, screening counts, exclusion reasons, and a flow diagram; these should be added only if the authors conduct and document a full systematic review process.

3.2 Source Selection Strategy

Sources were selected for conceptual relevance to four domains: youth football development, sports analytics platforms, data privacy and governance, and decentralized data ownership technologies. Priority was given to peer-reviewed literature, legal and regulatory materials, official platform documentation, and reputable industry reports. Commercial websites and platform materials were used only to understand current market practices and technological capabilities, not as standalone proof of contested academic claims.

3.3 Inclusion and Exclusion Criteria

The review included sources that addressed at least one of the following: youth athlete data, football analytics, biometric or performance tracking, algorithmic talent identification, data protection in sport, blockchain-based identity or data governance, or sports data markets. Sources were excluded when they were unrelated to youth sport, lacked relevance to data governance, provided purely promotional claims

without analytical value, or could not be connected to the article's ethical, legal, economic, or technical questions.

3.4 Analytical Procedure

The analysis followed a thematic coding logic. Source material was organized around recurring themes: data extraction, standardization, monetization, platform dependency, algorithmic bias, consent, privacy, cybersecurity risk, grassroots inequality, and potential governance solutions. These themes were then compared across stakeholder groups, including athletes, families, academies, platforms, professional clubs, regulators, and technology providers.

3.5 Case Selection and Ethical Treatment of Examples

The illustrative case discussions are not empirical case studies and do not involve private or confidential data. They rely exclusively on publicly available information from academic, legal, industry, platform, club, or media sources. Their purpose is to demonstrate broader governance tensions; they do not assert access to medical records, biometric datasets, internal platform data, commercial contracts, family decisions, or confidential club analytics.

3.6 Ethical Considerations

Because the study uses only publicly available secondary materials and does not collect personal data from participants, formal human-subjects ethics approval was not required. Nevertheless, the article follows principles of minimization, accuracy, non-harm, and responsible representation, especially when discussing young athletes and sensitive biometric or health-related information.

4. Conceptual Framework

4.1 Data Commodification Theory

Data commodification refers to the process by which information transforms from a public good or personal asset into a marketable commodity. In sports contexts, this involves the extraction, packaging, and monetization of athlete data through digital platforms. Drawing from critical political economy perspectives, this framework examines how data becomes alienated from its original context and repurposed for commercial gain (Zhou, 2025).

The theoretical foundation encompasses three key dimensions:

- Extraction The systematic collection of performance metrics, biometric data, and behavioral information from young athletes through various technological means.
- Standardization The transformation of disparate data formats into unified, comparable metrics that facilitate trading and analysis across different contexts.
- Monetization The conversion of standardized data into economic value through subscription models, licensing agreements, and marketplace transactions.

4.2 Ethical Framework for Youth Sports Data

The vulnerability of minor athletes necessitates a robust ethical framework that prioritizes protection over profit. This framework integrates four core principles:

- Autonomy Young athletes and their families must maintain control over their personal data, including the right to access, modify, and delete information.
- Beneficence Data collection and usage should primarily serve developmental goals rather than purely commercial interests.
- Justice Economic benefits derived from youth football data should be distributed fairly among all stakeholders, particularly the athletes who generate the value.
- Transparency All data practices must be clearly communicated to participants and families through accessible, age-appropriate explanations.

4.3 Technological Determinism vs. Social Construction

This research adopts a balanced perspective that acknowledges both technological determinism and social construction of technology. While advanced analytics tools create new possibilities for talent identification and development, their implementation and impact are shaped by social, economic, and cultural factors. This dual perspective enables analysis of how technical capabilities interact with institutional practices to produce specific outcomes in youth football data ecosystems.

5. Youth Football Development and the Data Revolution

5.1 Historical Evolution of Youth Football Analytics

The integration of data analytics in youth football has progressed through distinct phases:

- Phase 1 (Pre–2010): Manual Tracking Coaches relied on handwritten notes, basic statistics, and subjective observations to evaluate player performance. Data collection was inconsistent and primarily served immediate training purposes.
- Phase 2 (2010–2018): Digital Transition The introduction of video analysis software and basic performance tracking systems enabled more systematic data collection. Platforms like early Wyscout versions focused on professional football but began expanding to youth markets.
- Phase 3 (2019–2023): AI Integration Machine learning algorithms and computer vision technologies revolutionized data analysis, enabling automated performance assessment and predictive modeling. Wearable sensors became more accessible, allowing comprehensive biometric monitoring.
- Phase 4 (2024–Present): Ecosystem Integration Current systems integrate multiple data sources—video, sensors, psychological assessments—into unified platforms. Hudl Wyscout's Physical Data

pack exemplifies this trend, providing detailed running metrics and intensity analysis for U23 players across multiple leagues (Hudl, 2025b).

5.2 Impact on Talent Identification and Development

The data revolution has fundamentally altered talent identification processes. Traditional scouting, which relied heavily on subjective evaluation and limited exposure, has been supplemented—and in some cases replaced—by data-driven approaches. AI-powered platforms can now analyze individual player performance and biometric data to create customized training programs, enabling more precise talent identification and development planning (Kim, 2025).

However, this transformation raises critical questions about access and equity. Elite academies with substantial resources can implement comprehensive data collection systems, while grassroots organizations often lack the infrastructure and expertise to compete. This digital divide threatens to reinforce existing inequalities in youth football, potentially excluding talented players from under-resourced backgrounds.

5.3 International Example: La Masia Academy

Technology Infrastructure La Masia has implemented one of the most advanced youth development systems globally, utilizing the “Barça GPS Tracker” device across all academy levels (FC Barcelona, 2024). This comprehensive system collects real-time data on:

- Player speed and acceleration metrics
- Distance covered during training and matches
- Heart rate monitoring and recovery times
- Technical accuracy and skill execution

Performance Tracking Implementation The academy's data collection system enables coaches to make evidence-based promotion decisions. GPS trackers, heart rate monitors, and performance software collect real-time data on everything from player speed to recovery times, allowing for individualized development plans (SmartTech, 2025).

Data-Driven Decision Making La Masia's approach exemplifies how elite academies leverage technology for talent identification. The system collects a wide range of data crucial for comprehensive player development, enabling coaches to track progression and identify areas for improvement (Oliver Sports, 2025).

5.4 Regional Example: Algerian Youth Football Academies

Digital Divide Reality Algerian sports administration faces significant challenges in sports professionalism due to the digital divide (Benseghir & Benyoucef, 2025). Many youth academies in the

region still rely on manual tracking methods, creating a substantial gap compared to European counterparts.

Current Practices Local academies typically use:

- Paper-based performance recording sheets
- Excel spreadsheets for basic statistics
- Manual video recording without analytical software
- Subjective coach evaluations without data validation

Impact on Talent Identification This technological gap limits the ability of Algerian academies to compete internationally. Without access to advanced analytics platforms, talented players may be overlooked or undervalued in the global market.

Resource Constraints The digital divide in Algerian youth football reflects broader socioeconomic challenges, where limited funding prevents investment in modern tracking technologies and data analytics systems.

6. Technical Infrastructure of Digital Platforms

6.1 Overview: Hudl, Wyscout, and InStat

Hudl Wyscout Integration

Hudl's acquisition of Wyscout represents a landmark consolidation in football data analytics. The combined platform offers extensive video and statistical data on football matches, players, and teams, serving as a comprehensive repository for football analysis (Hudl, 2025a).

Users can navigate through performance and career data, reports, and videos using intuitive search and query tools, making it a dominant force in global talent scouting (Wyscout, 2025).

The integration of physical data packs allows for sophisticated analysis of young prospects, particularly U23 players across multiple leagues worldwide. This capability enables detailed assessment of running metrics, intensity levels, and physical development patterns, providing valuable insights for talent identification (Hudl, 2025b).

InStat Platform Capabilities

InStat delivers advanced video analysis capabilities, detailed statistical tracking, opponent scouting tools, and customizable performance metrics that cater to professional clubs and serious analysts (InStat, 2025).

The platform provides access to sophisticated statistics such as Expected Goals (xG), Expected Assists (xA), and progressive passes, which are crucial for modern football analysis.

Market Positioning and Differentiation

These platforms compete on several dimensions:

- Data Coverage: Number of leagues, matches, and players tracked
- Analytical Depth: Sophistication of metrics and predictive capabilities
- User Experience: Interface design and ease of use
- Integration Capabilities: Compatibility with other systems and data sources

6.2 Data Pipelines: From Pitch Sensors to Cloud Storage

Sensor Technologies and Data Collection

Modern football data collection relies on sophisticated sensor technologies that capture comprehensive performance metrics:

Wearable Sensors GPS trackers, accelerometers, and heart rate monitors provide real-time biometric data during training and matches. These devices can be unobtrusively worn on the body and are increasingly used in soccer training to collect data on players' physical performance (Kranzinger et al., 2025). Research shows that parameters obtained from wearable sensors are successful in forecasting the recovery conditions of youth football players, enabling more precise training load management (Teixeira et al., 2025).

Computer Vision Systems Advanced camera networks and AI-powered tracking systems automatically capture player movements, ball trajectories, and tactical patterns without requiring physical sensors. These systems can process vast amounts of visual data to generate detailed performance metrics.

Footwear-Mounted Sensors Playermaker sensors represent innovative technology, being the first FIFA-certified footwear-mounted solution for official matches, providing unique insights into foot-specific performance metrics (Playermaker, 2025).

Data Processing and Standardization

Raw sensor data undergoes extensive processing to transform it into usable analytics:

Real-Time Processing Data pipelines convert on-pitch events into structured signals—goals are timestamped, clocks advance, and players are tagged with relevant metadata (Zone14 AI, 2025). This enables immediate feedback and analysis during training sessions and matches.

Quality Assurance Automated validation algorithms check data integrity, identifying and correcting errors or anomalies in sensor readings.

Normalization Performance metrics are standardized across different contexts (leagues, competitions, equipment) to enable fair comparisons and meaningful analysis.

Cloud Infrastructure and Storage

Cloud platforms have revolutionized data storage and accessibility in football analytics. Teams can now store large volumes of footage, sensor data, and metadata without the need for complex on-site servers (Network World, 2025). This infrastructure enables:

Scalability Cloud storage solutions like Azure Blob Storage provide highly scalable and cost-effective object storage, perfect for handling large volumes of unstructured data like match footage (Microsoft Azure, 2025).

Accessibility Authorized users can access data from anywhere with internet connectivity, facilitating remote scouting and collaboration.

Security Advanced encryption and access control mechanisms protect sensitive athlete information from unauthorized access.

Analytics Integration Cloud platforms enable seamless integration with AI and machine learning tools, allowing real-time data analysis and predictive modeling.

6.3 AI and Machine Learning in Performance Prediction

Machine Learning Algorithms and Applications

AI-driven systems have transformed football prediction from guesswork to data-driven science. Modern AI sports prediction models typically achieve 65–75% accuracy across major leagues, significantly outperforming random guessing (API4AI, 2025). These systems employ various machine learning techniques:

Supervised Learning Algorithms trained on historical performance data can predict future outcomes, such as player market values based on on-field performance metrics (Moya et al., 2025).

Unsupervised Learning Clustering algorithms identify patterns and groupings in player data, revealing hidden talent pools or development trajectories.

Deep Learning Neural networks process complex, multi-dimensional data to identify subtle performance indicators that might escape human observation.

Player-Centric Prediction Models

Recent advances emphasize player-centric approaches that focus on individual development rather than team outcomes. These models analyze comprehensive datasets including physical metrics, technical skills, tactical understanding, and psychological factors to create personalized development pathways (Barraclough, 2025).

Research examining the integration of machine learning techniques in professional football focuses on two key areas: player and team performance analysis, and predictive modeling for talent identification (Moya et al., 2025). These applications enable more accurate assessment of young players' potential and optimal development strategies.

Ethical Considerations in AI Implementation

The deployment of AI in youth football raises critical ethical questions:

Algorithmic Bias AI models may perpetuate existing inequalities if training data reflects historical biases in access to quality coaching, facilities, or competition opportunities.

Transparency Complex machine learning algorithms can be “black boxes,” making it difficult to understand how predictions are generated and potentially undermining trust in the system.

Accountability When AI-driven decisions impact young athletes' careers, questions arise about who bears responsibility for errors or harmful outcomes.

Privacy Advanced analytics often require extensive personal data, raising concerns about surveillance and data protection for minors.

7. Economic Dimensions of Data Commodification

7.1 Value Chain Analysis

The youth football data value chain encompasses multiple stakeholders, each contributing to and benefiting from the commodification process:

Data Producers Young athletes generate the raw performance and biometric data that forms the foundation of the value chain. Despite being the primary source of value, they typically receive minimal direct compensation.

Data Collectors Local academies, clubs, and technology providers gather and initially process data. They invest in infrastructure and expertise but often lack the scale to monetize data effectively.

Data Aggregators Global platforms like Hudl Wyscout aggregate data from multiple sources, standardizing and enriching it to create comprehensive databases. These entities capture significant value through their scale and market position.

Data Consumers Professional clubs, agents, scouts, and researchers purchase access to aggregated data for talent identification, recruitment, and analysis purposes.

Technology Providers Companies developing sensors, software, and analytics tools benefit from the growing demand for data collection and analysis capabilities.

7.2 Revenue Models and Monetization Strategies

Digital platforms employ diverse revenue models to monetize youth football:

Subscription Services Tiered access models where users pay monthly or annual fees for varying levels of data access and analytical capabilities. Premium tiers often include advanced features like predictive analytics and custom reporting.

Data Licensing Selling aggregated, anonymized datasets to research institutions, equipment manufacturers, or betting companies. This model generates revenue while potentially protecting individual privacy.

Talent Marketplace Fees Commission-based models connecting players with opportunities, where platforms take a percentage of transfer fees or representation agreements facilitated through their systems.

Analytics Tools Premium features for advanced statistical modeling, video analysis, and performance prediction that require specialized expertise and computational resources.

Partnership and Sponsorship Collaborations with equipment manufacturers, sports brands, and media companies that leverage platform data for marketing, product development, or content creation.

7.3 Economic Distribution Patterns

Analysis of revenue distribution reveals significant inequalities in the youth football data ecosystem:

Platform Dominance Major platforms capture the largest share of economic value, benefiting from network effects, brand recognition, and economies of scale. Hudl's acquisition strategy demonstrates market consolidation trends that may further concentrate economic power (Hudl, 2025a).

Club Advantages Professional clubs with substantial resources can afford premium data subscriptions and in-house analytics capabilities, enhancing their talent identification and development processes while smaller organizations struggle to compete.

Academy Challenges Local academies often face a double bind—they need data analytics to remain competitive but lack the resources to implement sophisticated systems. Many become dependent on platform partnerships that may not provide fair value.

Player Exclusion Young athletes, despite being the source of valuable data, typically receive minimal direct financial benefit. Their compensation comes indirectly through improved development opportunities, but this varies widely based on academy resources and platform relationships.

7.4 Market Consolidation and Competition Concerns

The youth football data analytics market shows signs of consolidation, with major players acquiring specialized platforms to expand market dominance. This trend raises concerns about:

Reduced Competition Fewer independent platforms may lead to higher prices, reduced innovation, and limited choice for consumers.

Barriers to Entry New entrants face significant challenges competing with established platforms that have extensive data repositories, brand recognition, and customer relationships.

Dependency Risks Academies and clubs may become overly dependent on specific platforms, limiting their ability to negotiate favorable terms or switch providers.

Innovation Constraints Market concentration could stifle innovation if dominant platforms prioritize short-term profits over long-term technological advancement.

8. Ethical, Legal, and Cybersecurity Considerations

8.1 Privacy and Informed Consent

Legal Frameworks and Compliance Requirements

Youth sports organizations must navigate complex legal landscapes regarding data privacy and protection. GDPR compliance requires clubs working with youth players to receive consent from parents or guardians before processing data related to registration, health, or performance (GDPR Local, 2025). Similar regulations exist in other jurisdictions, creating a patchwork of requirements that organizations must navigate.

California Assembly Bill 734, effective January 1, 2025, represents emerging legislative attention to youth sports safety and data protection, potentially setting precedents for other regions (O'Connor's Law, 2025).

Informed Consent

Challenges

Obtaining truly informed consent from minors and their families presents significant challenges:

Complexity of Data Usage Modern analytics involve sophisticated processes that may be difficult for non-technical stakeholders to understand fully. Parents may consent to basic data collection without

comprehending how extensively their children's data might be used, shared, or monetized.

Power Imbalances Young athletes and families may feel pressured to consent to extensive data collection to maintain participation opportunities, limiting their ability to make truly voluntary decisions.

Long-term Implications Data collected during youth development may have consequences years later, but current consent processes rarely address these long-term implications adequately.

Dynamic Consent Models Emerging approaches emphasize ongoing, adaptable consent processes that allow participants to modify their preferences over time as understanding and circumstances evolve.

Best Practices for Privacy Protection

Organizations can implement several strategies to enhance privacy protection:

Data Minimization Collect only the data necessary for specific developmental purposes, avoiding excessive or unnecessary information gathering.

Anonymization and Pseudonymization Remove or mask personally identifiable information from datasets used for research or commercial purposes.

Transparent Communication Provide clear, accessible explanations of data collection practices, usage purposes, and sharing arrangements.

Regular Audits Conduct independent privacy assessments to ensure compliance with ethical standards and regulatory requirements.

8.2 Algorithmic Bias in Talent Identification

Sources of Algorithmic Bias

AI systems in sports analytics can perpetuate or amplify existing inequalities through several mechanisms:

Training Data Bias If historical data reflects systemic inequalities in access to quality coaching, facilities, or competition opportunities, algorithms trained on this data may replicate these biases.

Feature Selection Bias The choice of which performance metrics to prioritize can disadvantage players whose strengths do not align with conventional wisdom or available measurement tools.

Context Ignorance Algorithms may fail to account for contextual factors like socioeconomic background, access to resources, or cultural differences that affect performance metrics.

Feedback Loops Biased predictions can create self-reinforcing cycles where certain groups receive fewer opportunities, leading to less data and further biased predictions.

Mitigation Strategies

Addressing algorithmic bias requires multi-faceted approaches:

Diverse Training Data Ensure training datasets represent diverse populations and contexts to reduce bias in algorithmic predictions.

Bias Detection and Correction Implement systematic processes to identify and correct bias in algorithms, including regular audits and validation against real-world outcomes.

Human Oversight Maintain human involvement in critical decisions, particularly those affecting young athletes' careers, to provide context and judgment that algorithms may lack.

Transparency and Explainability Develop algorithms that can explain their predictions in understandable terms, enabling stakeholders to assess fairness and accuracy.

Stakeholder Involvement Include diverse perspectives in algorithm development and validation processes, particularly from underrepresented groups.

8.3 Security Protocols for Protecting Minors' Biometric Data

Biometric Data Risks

Collecting biometric data on minors creates particularly high stakes for privacy and security. Unlike passwords or other credentials, biometric information cannot be changed if compromised, creating permanent vulnerability (Spond, 2025).

A publicly reported lawsuit regarding biometric data collection in youth sports highlights the legal and ethical risks associated with this practice (Regulatory Review, 2025).

Comprehensive Security Framework

Organizations handling youth biometric data should implement robust security protocols:

Encryption Standards Use industry-leading encryption for data at rest and in transit, ensuring that even if data is intercepted, it remains unreadable.

Access Controls Implement strict role-based access controls, limiting data access to authorized personnel with legitimate needs.

Data Minimization Collect only essential biometric data and retain it for the minimum necessary period.

Incident Response Plans Develop comprehensive protocols for responding to data breaches, including notification procedures and mitigation strategies.

Third-party Audits Engage independent security experts to assess and validate security measures regularly.

Emerging Technologies for Enhanced Security

Blockchain technology offers promising solutions for securing athlete data. Blockchain provides a decentralized, secure, and tamper-proof method to manage athlete performance data, ensuring integrity and privacy (Bucea-Manea-Țoniș et al., 2025).

Decentralized identity management systems can verify athletes, coaches, and participants while maintaining secure credential verification processes (Principe et al., 2025).

Smart contracts on blockchain platforms can automate data sharing agreements, ensuring that terms are transparently enforced and that athletes receive fair compensation for their data usage (Bucea-Manea-Țoniș et al., 2025).

8.4 Vulnerability Analysis

Unauthorized Access Threats

Scenario A major breach of Wyscout's database could expose sensitive biometric data of thousands of youth players.

Potential Exploitation

- Insurance Industry: Biometric data (speed, heart rate, injury history) could be sold to insurance companies for risk assessment and premium calculation
- Betting Markets: Performance metrics and injury predictions could be exploited by betting syndicates for unfair advantages
- Competitive Intelligence: Rival clubs could access proprietary training methodologies and player development strategies
- Identity Theft: Personal information of minors could be used for fraudulent activities

Real-World Precedent In 2025, a Swedish sports platform suffered a massive data leak compromising the personal data of more than 2.1 million people, including names, contact details, identification numbers, and potentially sensitive biometric information (Spond, 2025).

Impact Assessment

- Financial: Estimated losses exceeding $50 million in platform value and legal liabilities
- Reputational: Loss of trust from clubs, players, and families
- Operational: Disruption of scouting and development activities
- Legal: Potential GDPR violations and regulatory fines

Data Integrity and Spoofing Risks

Data Spoofing Definition The deliberate manipulation of performance metrics to artificially enhance a player's market value or hide deficiencies.

Attack Vectors

- Sensor Tampering: Physical manipulation of wearable devices to record false performance data
- Software Exploitation: Hacking into data collection systems to modify recorded metrics
- Video Manipulation: Using deepfake technology to create misleading performance footage
- Algorithm Poisoning: Introducing biased data to corrupt machine learning models

Economic Consequences

- Transfer Market Manipulation: Artificially inflated player values leading to poor investment decisions
- Scouting Inefficiency: Wasted resources on players with manipulated profiles
- Competitive Disadvantage: Clubs making decisions based on compromised data
- Market Instability: Erosion of trust in data-driven talent identification

Insider Threats and Data Misuse

Risk Categories

- Employee Misconduct: Staff with access privileges selling data to unauthorized parties
- Third-Party Exploitation: Partners or vendors misusing shared data for commercial gain
- Coach Manipulation: Influencing data collection to favor certain players
- Family Exploitation: Parents or guardians pressuring players to manipulate performance data

Mitigation Strategies

- Access Controls: Implementing strict role-based permissions and least privilege principles
- Audit Trails: Comprehensive logging of all data access and modification activities
- Employee Training: Regular cybersecurity awareness programs and ethical guidelines
- Contract Enforcement: Legal agreements with clear data usage restrictions and penalties

9. Findings and Discussion

This section synthesizes the article’s findings around four interconnected patterns: commodification, unequal access, illustrative athlete data risks, and the potential of a decentralized governance model. The findings should be understood as conceptual and analytical conclusions derived from secondary-source analysis rather than statistical results from primary data collection.

9.1 Patterns of Commodification

Standardization and Aggregation Processes

The research reveals systematic patterns in how youth football data transforms from local metrics to global commodities. At grassroots levels, data collection begins with basic performance metrics—goals scored, minutes played, physical measurements—often recorded manually by coaches.

As players progress through academy systems, data becomes increasingly sophisticated, incorporating video analysis, biomechanical measurements, and psychological assessments.

Digital platforms facilitate this transformation by standardizing disparate data formats into unified profiles. Hudl Wyscout aggregates performance data from over 600 leagues worldwide, creating comprehensive player databases that transcend local contexts (Hudl, 2025a).

This standardization enables data to function as a tradable commodity, valued for its predictive potential rather than its original developmental purpose.

Value Extraction Mechanisms

Analysis of platform operations reveals diverse monetization strategies that extract value from youth football:

Subscription Tiering Platforms create artificial scarcity by restricting access to premium features, compelling users to pay for comprehensive data access.

Data Enrichment Raw performance data is enhanced with contextual information, predictive analytics, and comparative benchmarks, increasing its market value.

Network Effects Platforms leverage their user base to create self-reinforcing value propositions—more users generate more data, which attracts more users and increases platform value.

Integration Lock-in Comprehensive platform ecosystems make it difficult for users to switch providers, reducing competition and enabling price increases.

Stakeholder Power Dynamics

The commodification process creates significant power imbalances among stakeholders:

Platform Dominance Major platforms control data infrastructure, standards, and market access, enabling them to capture disproportionate value.

Academy Dependency Local academies become dependent on platform services for talent identification and development, limiting their bargaining power.

Player Vulnerability Young athletes lack resources, expertise, and legal standing to negotiate favorable data terms, leaving them exposed to exploitation.

Regulatory Gaps Existing legal frameworks inadequately address the unique challenges of youth sports data commodification, creating opportunities for abuse.

9.2 The Digital Divide between Grassroots and Elite Academies

Resource Disparities

The research identifies stark disparities in data analytics capabilities between elite and grassroots academies:

Financial Barriers Premium platform subscriptions, sensor equipment, and technical expertise require substantial investment that many local academies cannot afford.

Technical Expertise Implementing and utilizing advanced analytics requires specialized knowledge that may be unavailable in resource-constrained settings.

Data Quality Elite academies can afford high-quality sensors and comprehensive data collection systems, while grassroots organizations often rely on basic metrics or manual tracking.

Integration Capabilities Well-resourced academies can integrate multiple data sources and analytics tools, creating holistic player profiles, while smaller organizations struggle with fragmented systems.

Impact on Talent Identification

These disparities have significant implications for talent identification and development:

Systematic Exclusion Talented players from under-resourced backgrounds may be overlooked if their performance is not captured by sophisticated analytics systems.

Reinforced Inequalities Data-driven talent identification may perpetuate existing socioeconomic disparities by favoring players with access to advanced training and technology.

Limited Mobility Players from grassroots academies may face barriers transitioning to elite systems if their data is not compatible or comprehensive enough.

Innovation Constraints Resource limitations prevent many local academies from experimenting with new analytics approaches or adapting to evolving best practices.

9.3 Illustrative Case Discussions: Public Information, Athlete Welfare, and Market Value

The following examples are used only to illustrate general governance issues using publicly available information. They should not be read as claims about confidential medical records, private biometric datasets, platform-held data, commercial contracts, or internal club decision-making.

Illustrative Discussion 1: Ansu Fati - Public Injury Reporting, Data Governance, and Athlete Welfare

Ansu Fati is used here as an illustrative example because public reporting around his career has frequently connected elite youth development, injury management, and high-value player evaluation. This discussion does not assess his private medical condition and does not claim access to clinical records, internal FC Barcelona monitoring systems, or confidential recovery analytics.

Publicly described performance-monitoring systems in elite academies, including GPS and wearable technologies, show how clubs may collect movement and workload indicators for training management. In this article, such systems are discussed at the level of governance risk only, not as evidence of any specific private dataset relating to Fati.

Examples of data categories commonly discussed in football performance governance include:

- Training load and workload indicators
- Heart-rate or recovery-related indicators where lawfully collected
- Movement, acceleration, and running-volume indicators
- Injury-risk or recovery-support indicators when handled under appropriate medical and privacy safeguards

From a governance perspective, the core issue is not any specific record about Fati, but the broader possibility that health-adjacent performance data can influence player valuation, recruitment decisions, insurance risk, and club bargaining power. This makes consent, purpose limitation, retention rules, and access controls essential.

The cybersecurity concern is therefore systemic: if sensitive athlete data in any elite-football environment were breached or misused, it could expose minors or young players to reputational, financial, and welfare harms. The example is used to illustrate this risk, not to allege any breach or misuse involving Fati.

Ethically, the example highlights a general tension between player welfare and commercial value. Data can support safer training and rehabilitation, but it requires transparent governance to prevent health-adjacent information from being repurposed beyond the athlete's interests.

Illustrative Discussion 2: Lamine Yamal - Public Performance Visibility, Youth Stardom, and Data Value

Lamine Yamal is used here as an illustrative example of how a young player's public performance profile can acquire exceptional visibility in modern football. The discussion relies only on publicly available

information and does not claim access to private academy records, biometric data, family decisions, or commercial contracts.

Elite academy environments increasingly use performance analytics to support development planning. For this article, Yamal's case is treated as a governance example of how public match data, scouting attention, and youth visibility may interact; it is not presented as evidence of any specific private analytics process.

Data-informed development can help coaches manage exposure, training load, and progression, but in youth contexts it must be balanced against privacy, psychological pressure, and long-term welfare.

Market Commodification In public football data ecosystems, high-profile young players can become represented through quantifiable indicators and video-based evaluation. These indicators may include:

- Expected Goals (xG) and Expected Assists (xA) metrics
- Progressive passing statistics
- Defensive contribution metrics
- Physical development tracking

Economic and Ethical Impact The economic value attached to a public performance profile can increase rapidly when a young player attracts global attention. The governance question is whether the athlete and family have meaningful visibility, control, and fair benefit when data-derived value is created around the player's development.

Privacy Challenges As public attention grows, the boundary between legitimate performance analysis and intrusive datafication becomes more difficult to manage. This article therefore treats Yamal's case as an example of systemic governance risk, not as a claim about any private or confidential information.

10. Proposed Framework: A Decentralized Model for Data Ownership

The proposed model is a conceptual framework designed to redistribute control and value within youth football data ecosystems. It should not be presented as a completed or validated technical product. Its practical value depends on future prototyping, usability testing, legal review, cybersecurity auditing, and stakeholder acceptance.

Regulatory Challenges

Despite mentioning laws and regulations, facing massive global platforms like "Wyscout" and "Hudl" requires an extremely complex international legal strategy that may need further detail in research. These platforms possess enormous legal resources and the ability to influence regulatory policy formulation, which requires:

- Forming international legal alliances comprising experts from various jurisdictions

- Developing multi-level legal strategies including local, regional, and international protection
- Coordinating with regulatory bodies in different countries to ensure mutual compliance
- Preparing alternative plans to address potential legal challenges in different markets

Application Testing

A full evaluation of the proposed model would require preliminary results from "field studies" or "prototype testing" of the proposed decentralized system to ensure ease of use by non-technical player families. A future pilot application should include:

- Usability tests with families from diverse technical backgrounds
- Assessment of understanding of consent mechanisms and data management by non-technical users
- Measurement of user adoption and retention rates during the trial period
- Collection of qualitative feedback on user experience and interface improvements based on that feedback
- Verification of the effectiveness of automatic compensation mechanisms and the specific tools and methodologies for this phase are detailed in the following tables. Table 1 outlines usability testing tools, while Table 2 presents user satisfaction metrics. Interaction testing for decentralized technology is shown in Table 3. The pilot implementation roadmap is found in Table 4, and educational assessment tools in Table 5. Key Performance Indicators are defined in Table 6. Security awareness testing is detailed in Table 7, followed by continuous evaluation methods in Table 8. Recommended tools are listed in Table 9, and finally, the success criteria benchmarks are summarized in Table 10

Table 1
Usability Testing Tools

Tool/Method	Description	Target Metrics	Application Context
Direct Observation	Watching non-technical families use the system in real-time	Task completion time, error rates, confusion points	Lab or field settings with real families
Think-Aloud Protocol	Users verbalize thoughts while performing tasks	Cognitive load, understanding gaps, decision-making process	Individual testing sessions
Post-Use Interviews	Qualitative feedback collection after system interaction	Satisfaction levels, pain points, suggestions	After each testing session
Remote Usability Testing	Cloud-based testing with screen recording	Geographic diversity, accessibility issues	Distributed user groups

Table 2
User Satisfaction Measurement Tools

Tool	Type	Scale/Format	Key Dimensions Measured
System Usability Scale (SUS)	Quantitative	10-item questionnaire, 1–5 scale	Overall usability, learnability
User Experience Questionnaire (UEQ)	Quantitative	26 items, 7-point semantic differential	Attractiveness, perspicuity, efficiency, dependability, stimulation, novelty

Table 3
Decentralized Technology Interaction Testing

Testing Area	Specific Tools/Methods	Evaluation Criteria	Non-Technical Focus
Self-Sovereign Identity (SSI) Wallet	Prototype wallet testing	Ease of setup, backup recovery, daily use	Simplified onboarding, visual cues
Consent & Control Mechanisms	Scenario-based testing	Understanding of permissions, revocation	Clear language, visual indicators
Data Portability	Transfer simulation exercises	Success rate, time to complete, confidence	Step-by-step guidance, progress indicators
Blockchain Interaction	Simplified interface testing	Transaction understanding, fee awareness	Hidden complexity, plain language explanations

Table 4
Pilot Implementation Methodology

Phase	Activities	Duration	Success Indicators
Pre-Pilot Preparation	Recruit diverse families, create test scenarios, develop support materials	2–4 weeks	Representative sample, clear protocols
Controlled Testing	Small group (10–15 families) in supervised environment	2–3 weeks	Baseline usability metrics, major issues identified
Field Testing	Real-world usage in selected academies with 50–100 families	4–8 weeks	Adoption rates, retention, real-world challenges
Extended Validation	Larger scale (200 + families) across multiple locations	8–12 weeks	Scalability validation, diverse user feedback

Table 5
Knowledge and Education Assessment Tools

Assessment Type	Tool/Method	Measurement Focus	Frequency
Pre/Post Knowledge Tests	Multiple-choice quizzes	Conceptual understanding improvement	Before and after training
Simulation Scenarios	Role-playing exercises	Decision-making capability	Mid-point and final evaluation
Self-Assessment Tools	Confidence rating scales	User-perceived competence	Weekly during pilot
Practical Demonstrations	Hands-on task completion	Actual skill application	End of training modules

Table 6
Key Performance Indicators (KPIs)

Category	KPI	Target Metric	Measurement Method
Adoption & Retention	User adoption rate	> 70% of invited families	Registration analytics
	User retention rate	> 80% after 3 months	Active user tracking
	Time to first meaningful action	< 15 minutes	Session recording
Task Performance	Task completion rate	> 90% success	Usability testing
	Average task time	< 5 minutes for core tasks	Timing measurements
	Error rate	< 10% per task	Error logging
Understanding	Concept comprehension score	> 80% on knowledge tests	Pre/post assessments
	Independent data management	> 75% success rate	Practical demonstrations
	Data control confidence	> 4/5 on confidence scale	Self-assessment surveys
Satisfaction	SUS Score	> 75 (above average)	Standard SUS questionnaire
	UEQ Score	Above benchmark in all dimensions	UEQ survey
	CSAT Score	> 4/5 average	Post-use surveys
	Customer Satisfaction (CSAT)	Quantitative	Single question, 1–5 scale
	Net Promoter Score (NPS)	Quantitative	Single question, 0–10 scale
	In-depth Interviews	Qualitative	Semi-structured format
	Focus Groups	Qualitative	Group discussion (6–8 participants)

Table 7
Security and Privacy Awareness Testing

Testing Area	Method	Evaluation Criteria	Non-Technical Adaptation
Phishing Recognition	Simulated phishing attempts	Detection rate, reporting behavior	Realistic but safe scenarios
Privacy Settings Understanding	Settings configuration test	Correct configuration rate	Visual, intuitive interface
Informed Consent	Consent scenario testing	Understanding of implications	Plain language, visual aids
Data Sharing Awareness	Sharing simulation exercises	Appropriate sharing decisions	Clear consequences explanation

Table 8
Continuous Evaluation Methodology

Method	Implementation	Benefits	Frequency
A/B Testing	Compare different interface versions	Data-driven optimization	Ongoing for new features
Iterative Improvement	Collect feedback → implement changes → retest	Continuous enhancement	Bi-weekly cycles
Scalability Testing	Gradually increase user numbers	Identify bottlenecks early	Monthly assessments
Longitudinal Studies	Track same users over extended period	Usage pattern evolution	Quarterly follow-ups

Table 9
Recommended Testing Tools and Platforms

Tool Category	Specific Tools	Cost	Best For
Usability Testing	UserTesting, Lookback, Hotjar	$25–100/user	Remote testing, screen recording
Survey & Feedback	SurveyMonkey, Typeform, Google Forms	Free–$50/month	Quantitative data collection
Analytics	Mixpanel, Amplitude, Google Analytics	Free–$100/month	Behavioral tracking
Prototyping	Figma, InVision, Adobe XD	Free–$50/month	Interface testing before development
Session Recording	FullStory, SessionStack, Mouseflow	$50–200/month	User behavior analysis
A/B Testing	Optimizely, VWO, Google Optimize	$50–500/month	Interface optimization

Table 10
Success Criteria Summary

Dimension	Minimum Acceptable	Target Goal	Excellence Benchmark
Usability (SUS)	68 (Industry Average)	75+	85+
Task Success Rate	80%	90%	95%+
User Satisfaction (CSAT)	3.5/5	4/5	4.5/5+
Knowledge Retention	60% improvement	80% improvement	90%+ improvement
Adoption Rate	50%	70%	85%+
Retention Rate (3 months)	60%	80%	90%+

Technical Architecture

- Blockchain Infrastructure: A permissioned blockchain network provides secure, tamper-proof data storage and transaction recording. This decentralized approach ensures that no single entity controls the entire system (Bucea-Manea-Țoniș et al., 2025).
- Decentralized Identity Management: Blockchain-based identity systems verify athletes, coaches, and participants while maintaining secure credential verification processes (Principe et al., 2025). Each athlete has a unique digital identity that controls access to their data.
- Smart Contract Implementation: Automated agreements govern data usage, compensation distribution, and access permissions. These contracts execute transparently and cannot be altered without consensus (Bucea-Manea-Țoniș et al., 2025).
- Data Storage Solutions: Off-chain storage (IPFS, cloud storage) handles large data files while blockchain records metadata and access permissions, optimizing performance and cost.
- API Ecosystem: Open, standardized APIs enable interoperability between different platforms and systems, preventing vendor lock-in and promoting innovation.

Economic Model

- Revenue Sharing Mechanism: The model implements a transparent revenue-sharing system where:
- 20–30% supports academy operations and development programs
- 5–10% allocates to governance and oversight functions
- Microtransactions: Small, frequent payments enable fair compensation for individual data uses rather than lump-sum arrangements that may undervalue ongoing contributions.
- Collective Bargaining: Athletes can form data cooperatives to negotiate better terms with platforms and increase their collective bargaining power.

- Development Investment: A portion of revenue automatically funds player development programs, ensuring that economic benefits directly support athletic growth.

Implementation Roadmap

- Phase 1: Foundation (Months 1–6)
- Develop technical specifications and governance framework
- Establish pilot partnerships with willing academies and platforms
- Create legal and regulatory compliance frameworks
- Design user interfaces and consent mechanisms
- Phase 2: Pilot Implementation (Months 7–18)
- Deploy blockchain infrastructure and smart contracts
- Onboard initial cohort of athletes and academies
- Test data collection, sharing, and monetization processes
- Gather feedback and refine system design
- Phase 3: Scaling (Months 19–36)
- Expand to additional regions and platforms
- Develop interoperability standards and APIs
- Implement governance mechanisms and oversight structures
- Establish sustainable funding models
- Phase 4: Ecosystem Development (Months 37+)
- Foster innovation through open platform approach
- Develop additional services and applications
- Expand to other youth sports contexts
- Continuously improve based on stakeholder feedback

Advantages Over Current Models

- Empowerment: Athletes gain control over their data and participate directly in value creation.
- Transparency: All transactions and agreements are visible and verifiable, reducing information asymmetries.
- Fairness: Automated revenue sharing ensures equitable distribution of economic benefits.
- Security: Decentralized architecture reduces single points of failure and enhances data protection.
- Innovation: Open standards and APIs encourage competition and technological advancement.
- Sustainability: Revenue sharing creates incentives for long-term athlete development rather than short-term exploitation.

Challenges and Mitigation Strategies

- Technical Complexity: Blockchain and smart contract implementation requires specialized expertise. Solution: Partner with experienced technology providers and develop comprehensive training programs.
- Regulatory Uncertainty: Evolving data protection laws may create compliance challenges. Solution: Engage legal experts and regulatory bodies early in development process.
- Adoption Barriers: Existing platforms may resist decentralized models that threaten their business models. Solution: Demonstrate clear benefits for all stakeholders and create transition pathways.
- Cost Considerations: Initial implementation costs may be substantial. Solution: Seek funding from impact investors, sports organizations, and government grants focused on youth development.
- User Education: Athletes and families need to understand complex data ownership concepts. Solution: Develop accessible educational materials and support systems.

11. Limitations

- The article is based on secondary sources and does not include interviews, surveys, ethnographic observation, or fieldwork with athletes, families, academies, or platform operators.
- The proposed decentralized model has not yet been implemented or tested through a working prototype, pilot study, or field deployment.
- Some platform practices are not publicly transparent; therefore, the analysis may be limited by the availability and reliability of public documentation.
- Industry reports and commercial platform materials are useful for understanding market practices but may contain promotional bias and should be interpreted cautiously.
- The illustrative athlete discussions rely on public information only and cannot verify confidential medical, contractual, or platform-held data.
- The article does not present a formal PRISMA systematic review. If the manuscript is submitted as a systematic review, a complete PRISMA search log, screening table, and flow diagram must be added.

12. Conclusion and Recommendations

12.1 Summary of Key Findings

- Youth athlete data is transformed into commercial value through extraction, standardization, aggregation, enrichment, and restricted platform access.
- Elite academies and global platforms are positioned to capture disproportionate value because they control infrastructure, expertise, standards, and market access.
- Grassroots academies face financial and technical barriers that may reduce the visibility of talented players from under-resourced contexts.

- Minors and their families often have limited bargaining power, creating ethical concerns around consent, data ownership, long-term retention, and commercial monetization.
- AI-assisted talent identification can improve performance evaluation, but it may also reproduce inequality if data inputs and algorithmic assumptions are biased.
- A decentralized data ownership framework may improve transparency and revenue sharing, but it requires empirical validation, legal compatibility, usability testing, and cybersecurity review before adoption.

12.2 Practical Recommendations

For Local Academies

Implement Transparent Data Practices

- Develop clear policies explaining data collection purposes, usage, and sharing arrangements.
- Obtain informed consent from athletes and families using accessible language and visual aids.
- Provide regular updates on how data is used and what benefits accrue to participants.

Negotiate Fair Platform Agreements

- Form cooperatives with other local academies to increase bargaining power.
- Demand transparent pricing and revenue-sharing arrangements.
- Ensure data portability and interoperability in all platform contracts.
- Retain ownership rights to athlete data generated within the academy.

Invest in Ethical Infrastructure

- Prioritize data security and privacy protection in technology investments.
- Train staff on ethical data practices and athlete protection.
- Implement regular audits and compliance checks.
- Create internal ethics committees with stakeholder representation.

Develop Value-Added Services

- Offer data analysis consulting to other academies or clubs.
- Create educational content and training programs based on data insights.
- Host data-driven tournaments and showcase events.
- Partner with equipment manufacturers for sponsored programs.

For Digital Platforms

Adopt Ethical Design Principles

- Implement privacy-by-design approaches that protect athlete data by default.
- Create transparent consent processes that genuinely inform young athletes and families.
- Develop bias detection and correction mechanisms for AI algorithms.
- Establish independent ethics review boards to oversee data practices.

Create Fair Compensation Models

- Share economic value with data producers through transparent revenue-sharing arrangements.
- Implement microtransaction systems that compensate athletes for individual data uses.
- Support development programs that benefit young athletes directly.
- Provide free or discounted access to underserved communities.

Enhance Accessibility

- Develop tiered pricing models that accommodate smaller academies and grassroots organizations.
- Create open APIs and interoperability standards to reduce lock-in effects.
- Offer training and support programs for organizations with limited technical expertise.

Invest in user-friendly interfaces that make data analytics accessible to non-technical users.

Foster Innovation

- Support research on ethical data practices and athlete protection.

Collaborate with academic institutions and regulatory bodies.

- Participate in industry standards development.
- Share best practices and lessons learned with the broader community.

For Young Athletes and Families

Educate Yourself on Data Rights

- Seek comprehensive information about data collection practices before participation.
- Understand rights regarding data access, correction, and deletion.
- Ask questions about how data will be used, shared, and monetized.
- Negotiate data usage terms when possible, particularly for high-value prospects.

Maintain Data Control

- Keep records of consent agreements and data sharing arrangements.

- Regularly review and update data preferences.
- Request access to personal data and verify its accuracy.
- Exercise rights to data portability when changing academies or platforms.

Advocate for Fair Practices

- Join athlete associations or data cooperatives to increase collective bargaining power.
- Participate in governance processes for platforms and academies.
- Support regulatory initiatives that protect athlete rights.
- Share experiences and insights to inform policy development.

12.3 Policy Recommendations

Regulatory Frameworks

Youth Sports Data Protection Act

- Develop comprehensive legislation specifically addressing data collection, usage, and ownership rights for minors in sports contexts. Key provisions should include:

Mandatory transparency requirements for data practices.

Strict consent standards with parental involvement.

Data minimization and purpose limitation principles.

Right to data access, correction, and deletion.

Prohibition of exploitative data practices.

Independent Oversight Bodies

- Establish regulatory agencies with expertise in both sports and data protection to monitor compliance, investigate violations, and develop best practices. Include representation from athletes, parents, coaches, and technical experts.

Data Portability Standards

- Mandate standardized data formats and interoperability requirements to ensure young athletes can transfer their data between platforms and institutions without penalty or loss of information.

Algorithmic Accountability

- Require transparency and explainability for AI systems used in talent identification, including regular audits for bias and fairness.

Industry Self-Regulation

Ethical Guidelines Development

- Collaborative creation of industry standards for youth sports data practices through multi-stakeholder processes involving athletes, families, academies, platforms, and regulators.

Certification Programs

- Third-party verification of platforms and academies meeting ethical data standards, with regular audits and public reporting of compliance status.

Stakeholder Representation

- Include athlete advocates, parent representatives, and ethics experts in governance structures for platforms and industry associations.

Transparency Reporting

- Annual publication of data practices, revenue distribution, and impact assessments to enable public scrutiny and accountability.

12.4 Future Research Directions

- Longitudinal studies should track how early data collection affects athletic development, career trajectories, privacy expectations, and life outcomes over time.
- Cross-cultural comparisons should examine how regulatory environments and cultural attitudes shape youth sports data practices across regions.
- Algorithmic bias research should test whether AI talent identification systems produce unequal outcomes across socioeconomic, geographic, racial, gender, and academy-resource contexts.
- Economic modeling should estimate fair value distribution mechanisms that protect athletes while maintaining platform sustainability.
- Prototype testing should evaluate whether families and athletes can understand and use decentralized identity, consent, and revenue-sharing tools in real-world settings.
- Regulatory effectiveness studies should assess how different policy approaches influence athlete protection, innovation, market competition, and grassroots development.

12.5 Final Conclusion

The commodification of youth football data represents both unprecedented opportunity and significant challenge. Digital technologies offer capabilities for talent identification, development planning, and performance optimization that were unimaginable a decade ago.

However, these benefits must be balanced against ethical imperatives to protect vulnerable young athletes and ensure that technological advancement serves human development rather than purely commercial interests.

The proposed decentralized model for data ownership offers a promising pathway toward more equitable outcomes. By shifting control to athletes and implementing transparent, fair economic arrangements, this approach can harness the power of data analytics while safeguarding the rights and well-being of young participants.

Success requires collaboration among all stakeholders—academies, platforms, athletes, families, regulators, and researchers—to create systems that prioritize development over exploitation.

As the youth sports industry continues to mature and digital platforms expand their reach, the choices made today will shape the future of athlete development for generations to come. This research provides foundational understanding and practical guidance for that essential work, contributing to a more equitable and sustainable future for youth football development.

The central claim of this revised manuscript is therefore cautious but constructive: data-driven youth football can support development and opportunity, but only if governance mechanisms ensure transparency, informed consent, data portability, cybersecurity, algorithmic accountability, and fair distribution of value. Digital Athlete Sovereignty offers a conceptual foundation for that governance agenda.

Declarations

Funding

The author(s) declare that no funds, grants, or other financial support were received during the preparation of this manuscript.

Competing Interests

The author(s) declare no competing interests.

Ethics Approval

Ethics approval was not required because the manuscript is based solely on publicly available secondary sources and does not involve human participants, private records, interviews, surveys, experiments, or direct collection of personal data.

Consent to Participate and Consent to Publish

Not applicable.

Data Availability

No new datasets were generated or analyzed during the current study. The manuscript relies on publicly available academic, legal, industry, and platform sources cited in the reference list.

Author Contributions

The author(s) contributed to conceptualization, source review, analysis, manuscript drafting, and revision. This statement should be adjusted to match the final author list before journal submission.

Acknowledgements

Not applicable unless the author(s) wish to acknowledge institutional, supervisory, or editorial support.

References

1. API4AI (2025) Top AI trends transforming sports in 2025. Medium. https://medium.com/@API4AI/top-ai-trends-in-the-sports-industry-for-2025-98758bb89378
2. Barraclough S (2025) The implementation of a multidisciplinary athlete profiling tool. Front Psychol 16:1636386. https://doi.org/10.3389/fpsyg.2025.1636386
3. Bucea-Manea-Țoniș R et al (2025) Blockchain in Sports: A Comparative Analysis of Applications. Appl Sci 15(12):6829. https://doi.org/10.3390/app15126829
4. Cahill G, MacNamara Á (2025) Ethical paradoxes of data-informed talent identification in elite youth football academies: A Foucauldian analysis. University of Hull Repository
5. eHealth4Everyone (2025), October 29 How SportAPI AI drives data-driven football analytics. https://ehealth4everyone.com/?d-news-233785-2025-10-29-football-stats-company-how-sportapi-ai-drives-data-driven-football-analytics
6. eHealth4Everyone (2025), October 29 Live football data: Real-time metrics for fans, teams, and analysts. https://ehealth4everyone.com/?d-news-239799-2025-10-29-live-football-data
7. FC Barcelona (2024) Barça GPS Tracker launched for La Masía and global academies. https://www.fcbarcelona.com/en/football/first-team/news/2024-11/barca-gps-tracker-launched
8. FC Barcelona (2025) Lamine Yamal named LALIGA EA SPORTS Best Under-23 Player 2024/2025. https://www.fcbarcelona.com/en/football/first-team/news/2025-05/lamine-yamal-laliga-best-under-23-player
9. GDPR Local (2025), September 30 Protecting personal What football clubs should know. https://gdprlocal.com/football-clubs-gdpr/

10. Hudl (2025a) Football data analytics software • Hudl Wyscout. https://www.hudl.com/en_gb/products/wyscout
11. Hudl (2025b) Discovering top young prospects with Hudl Physical Data. https://www.hudl.com/blog/discovering-young-prospects-wyscout-physical-data
12. Hudl (2025c) Scouting Serie A: Physical Data standouts 2024/25. https://www.hudl.com/blog/seriea-physical-data-standouts-2024-25
13. Hudl (2024) How wearable technology elevates elite football. https://www.hudl.com/blog/wearable-technology-in-sport
14. InStat (2025) InStat football analytics platform. https://instatfootball.com/
15. Intel Market Research (2025) Global football software market outlook 2025–2032. https://www.intelmarketresearch.com/Global-Football-Analysis-922
16. Kim JH (2025) Ethical implications of artificial intelligence in sport. J Sport Health Sci 14(3):25000250. https://doi.org/10.1016/j.jshs.2025.03.005
17. Kranzinger S et al (2025) Performance tracking in female youth soccer through wearable sensors. Front Sports Act Living 7:1627820. https://doi.org/10.3389/fspor.2025.1627820
18. Li X (2024) Factors affecting the protection of data rights in sports events. Sci Rep 14:56074. https://doi.org/10.1038/s41598-024-56074-6
19. Market Growth Reports (2026) Youth sports software market report | Global insights [2035]. https://www.marketgrowthreports.com/market-reports/youth-sports-software-market-113852
20. Microsoft Azure (2025) Azure Blob Storage documentation. https://learn.microsoft.com/en-us/azure/storage/blobs/
21. Moya D et al (2025) Machine Learning Applied to Professional Football. Technologies 7(3):85. https://doi.org/10.3390/technologies7030085
22. Network World (2025), November 24 NFL, AWS drive football modernization with cloud, AI. https://www.networkworld.com/article/4095370/nfl-aws-drive-football-modernization-with-cloud-ai.html
23. N3XT Sports (2025) Sport's digital & data transformation guidebook 2024–2025. https://www.n3xtsports.com/report-data-digital-transformation-guidebook-2025/
24. O'Connor's Law (2025), February 2 Data protection changes are on the way for sports clubs and businesses. https://www.oconnors.law/news-views/data-protection-changes-are-on-the-way-for-sports-clubs-and-businesses/
25. Oliver Sports (2025) Oliver Sports performance tracking technology. https://www.oliversports.com/
26. Orrick (2025), April 9 Data privacy in sports: Key takeaways. https://www.orrick.com/en/insights/2025/04/data-privacy-in-sports-key-takeaways
27. Playermaker (2025) Playermaker football performance sensors. https://www.playermaker.com/
28. Principe V et al (2025) Transformative blockchain technological approaches to sports. PMC 12206727. https://doi.org/10.1101/2025.01.15.12206727

29. Project Play (2025) State of play 2025: 10 youth sports trends to watch. https://projectplay.org/state-of-play-2025/10-youth-sports-trends-to-watch
30. Pulse Ghana (2025), May 13 The economics behind youth football academies. https://www.pulse.com.gh/story/the-economics-behind-youth-football-academies-2025051309485149859
31. Regulatory Review (2025), August 12 NCAA athletes' right to privacy in the digital age. https://www.theregreview.org/2025/08/12/taylor-ncaa-athletes-right-to-privacy-in-the-digital-age/
32. SmartTech (2025), December 13 Best 10 wearable sensors for football. https://www.smarttech.fi/wearable-sensors-for-football/
33. Spond (2025), August 18 Protecting youth sports data and privacy. https://www.spond.com/en-us/news-and-blog/protecting-youth-sports-data/
34. Teixeira JE et al (2025) Player Tracking Data and Psychophysiological Features for Recovery Forecasting in Youth Football. Appl Sci 15(7):3718. https://doi.org/10.3390/app15073718
35. The Esk (2025), August 9 The analysis series: Premier League academy productivity and profitability 2015–2025. https://theesk.org/2025/08/09/the-analysis-series-premier-league-academy-productivity-and-profitability-2015-2025/
36. WilmerHale (2025), March 17 Sports and gaming law 2024 year in review: Top five developments. https://www.wilmerhale.com/en/insights/client-alerts/20250314-sports-and-gaming-law-2024-year-in-review-top-five-developments/
37. Wyscout (2025) Wyscout - Apps on Google Play. https://play.google.com/store/apps/details?id=com.wyscout.standard
38. Youth Sports Business Report (2025), December 19 Youth sports business in 2025: The year the industry grew up. https://youthsportsbusinessreport.com/youth-sports-business-in-2025-the-year-the-industry-grew-up/
39. Yu W et al (2025) International youth football research developments. ScienceDirect. https://doi.org/10.1016/j.sbspro.2025.01.001
40. Zone14 AI (2025), March 25 Advancements in football through AI, data analytics, and tracking technologies. https://zone14.ai/en/blog/football-through-ai-data-analysis-and-tracking-technologies/
41. Zhou D (2025) Artificial intelligence in sport: A narrative review of ethical and privacy concerns. J Sports Sci 43(8):2518694. https://doi.org/10.1080/02640414.2025.2518694

Figures

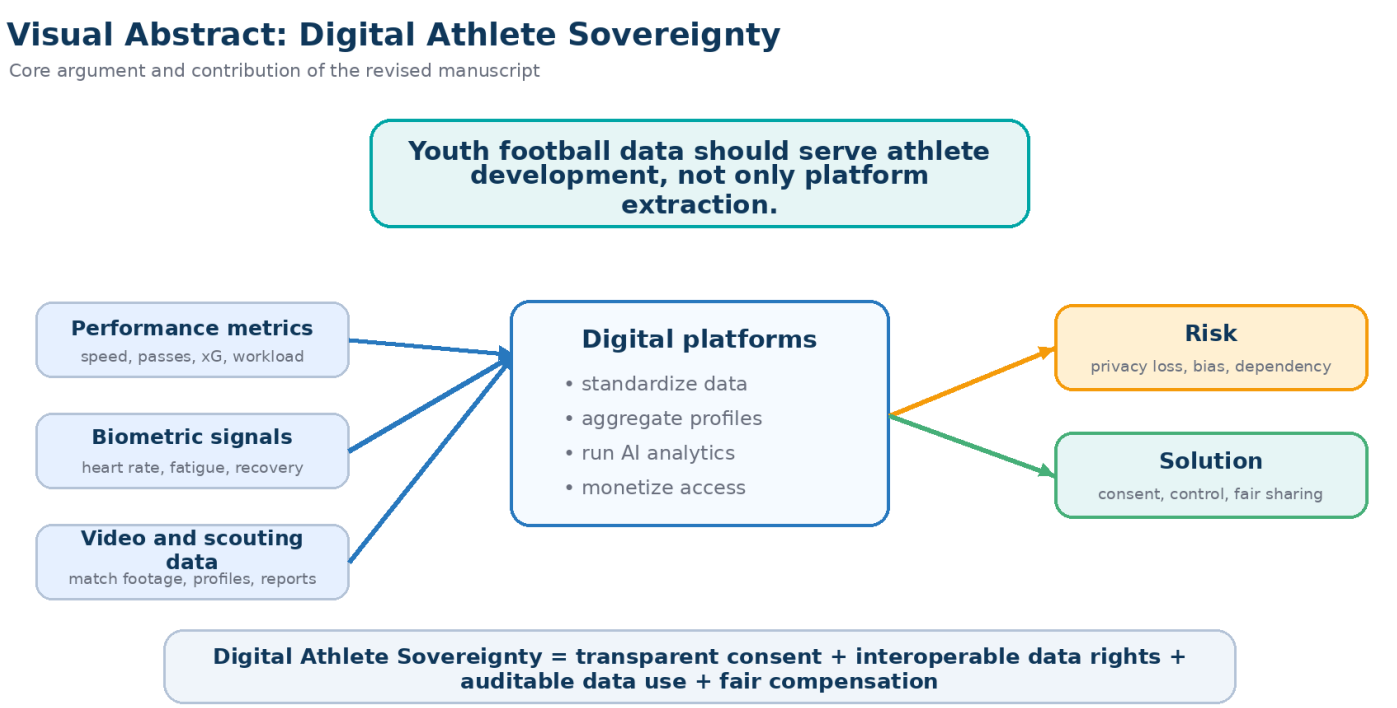


Figure 1

Visual abstract showing the article's core argument: youth football data should be governed through Digital Athlete Sovereignty rather than simple platform extraction.

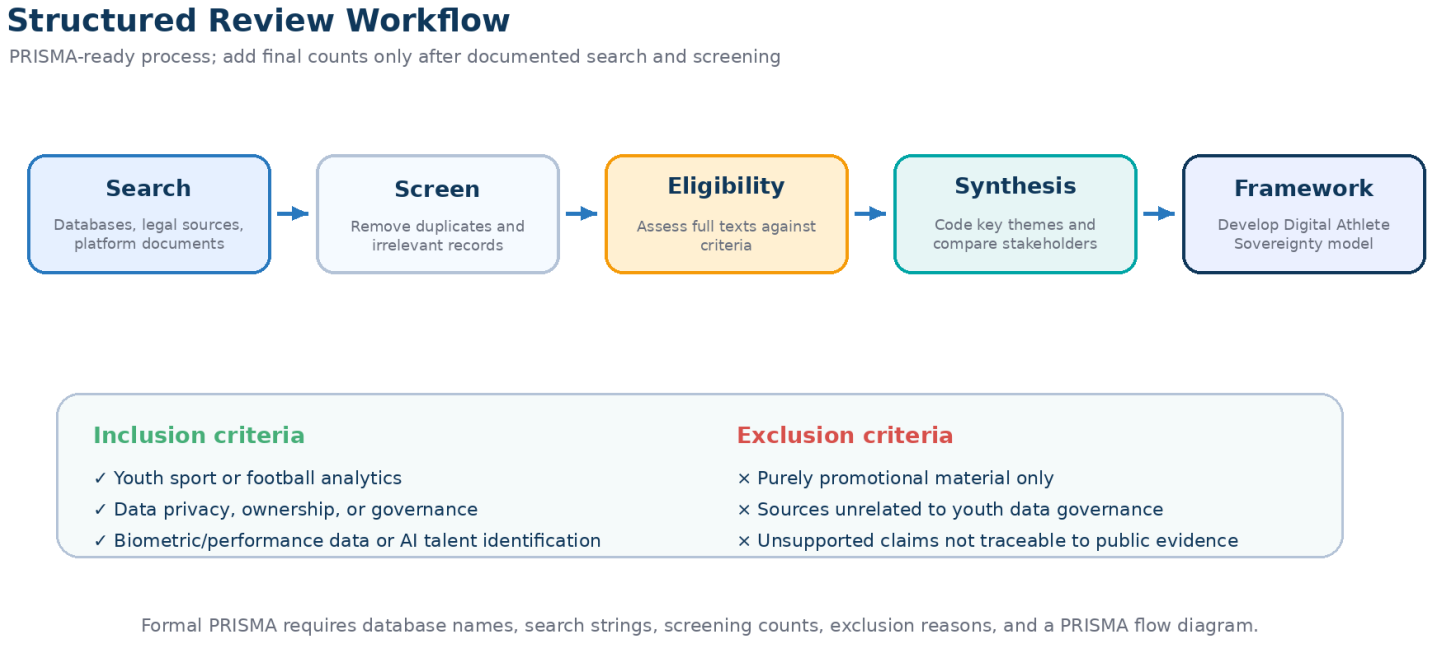


Figure 2

Structured secondary-source review workflow. The workflow is PRISMA-ready, but formal PRISMA counts should be added only after a documented database search and screening process.

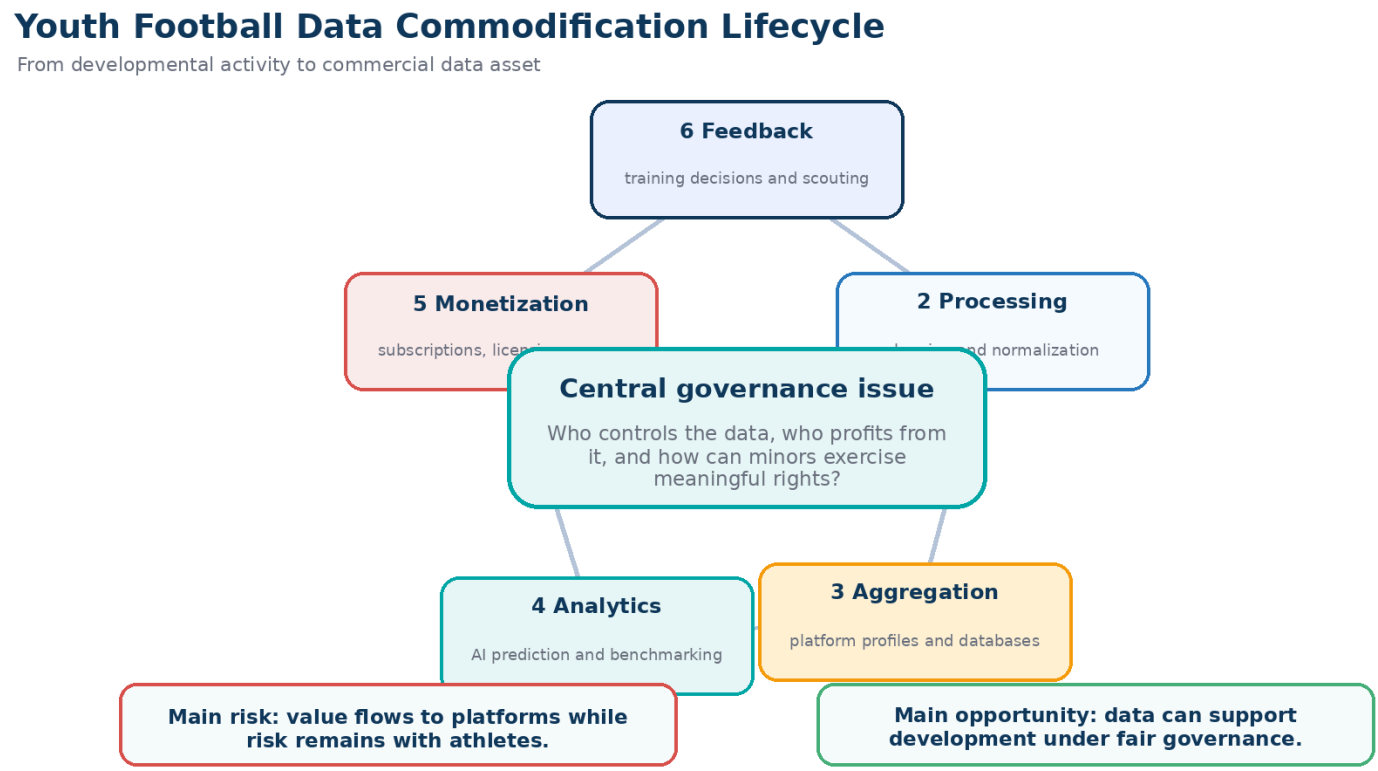


Figure 3

Youth football data commodification lifecycle from collection to processing, aggregation, analytics, monetization, and feedback into talent development.

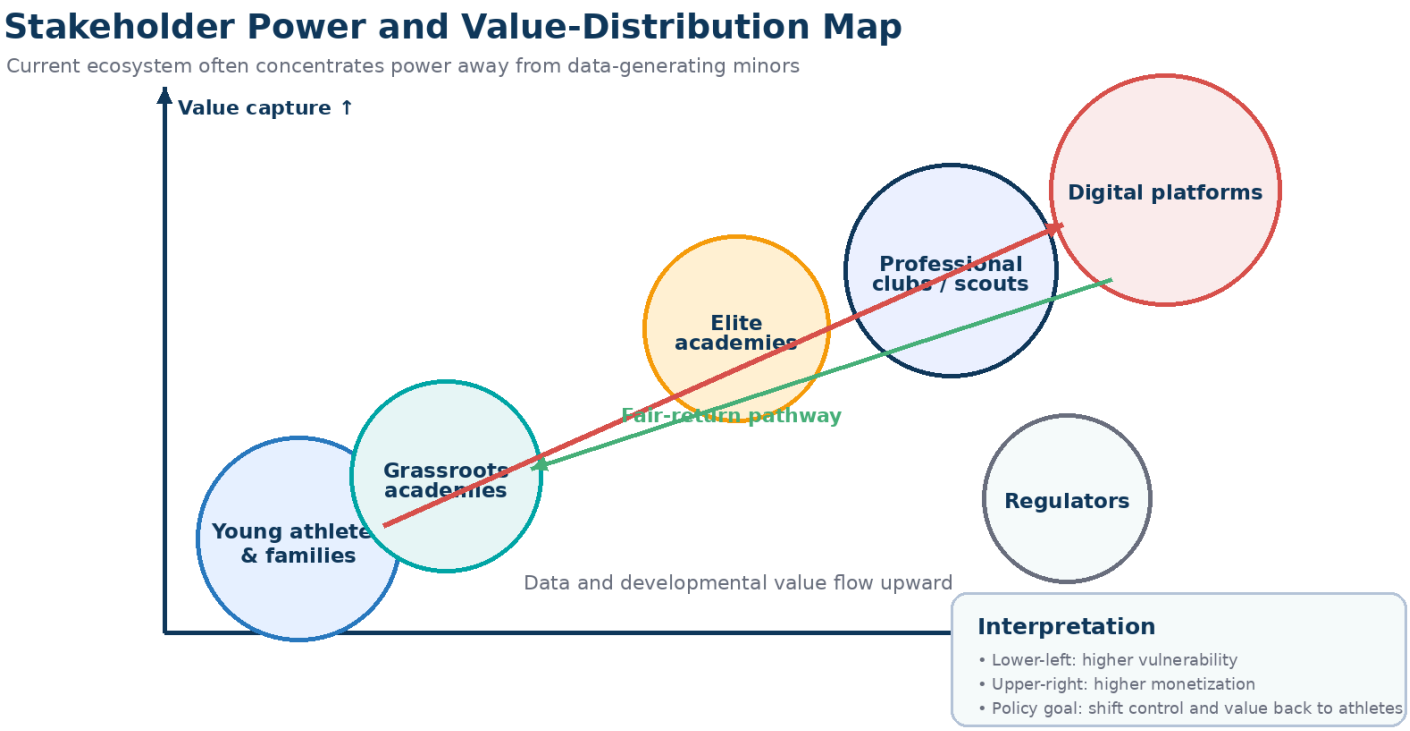


Figure 4

Stakeholder power and value-distribution map showing how data-generating minors often hold less bargaining power than platforms and elite market actors.

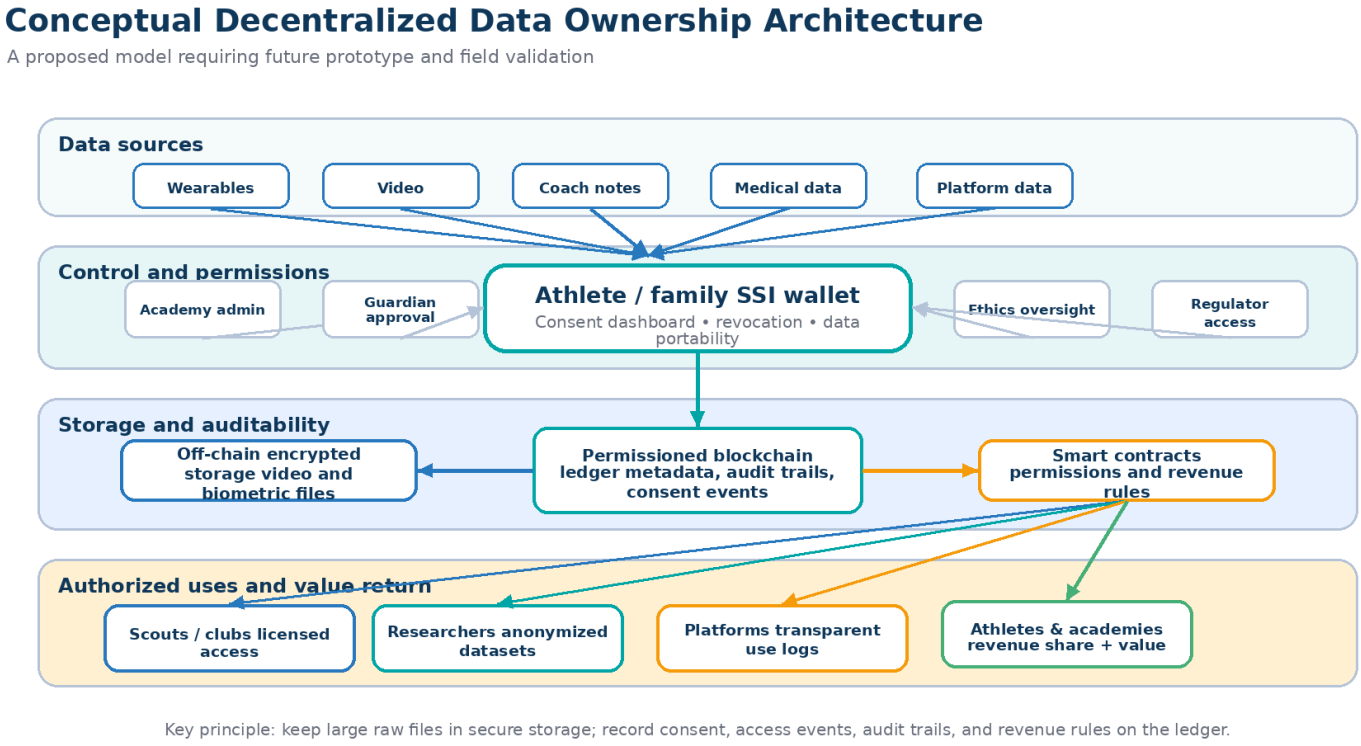


Figure 5

Conceptual decentralized data ownership architecture for Digital Athlete Sovereignty, combining SSI wallets, consent dashboards, secure storage, blockchain audit trails, smart contracts, and fair value return.