BMJ Open Sport & Exercise Medicine

Big data. Big potential. Big problems?

Stephen W West ,[1,2,3] Jo Clubb ,[4] Tracy A Blake ,[5] James Fern ,[6] Harry Bowles ,[6] Torstein Dalen-Lorentsen [7,8]

To cite: West SW, Clubb J, Blake TA, *et al*. Big data. Big potential. Big problems? *BMJ Open Sport & Exercise Medicine* 2024;**10**:e001994. doi:10.1136/bmjsem-2024-001994

Accepted 6 April 2024

[1]Centre for Health, and Injury & Illness Prevention in Sport, University of Bath, Bath, UK
[2]UK Collaborating Centre on Injury and Illness Prevention in Sport, University of Bath, Bath, UK
[3]Sport Injury Prevention Research Centre, Faculty of Kinesiology, University of Calgary, Calgary, Alberta, Canada
[4]Global Performance Insights, London, UK
[5]Faculty of Health Sciences, School of Rehabilitation Science, McMaster University, Hamilton, Ontario, Canada
[6]Department for Health, University of Bath, Bath, UK
[7]Department of Health Research, SINTEF Digital, Oslo, Norway
[8]Oslo Sports Trauma Research Center, Department of Sports Medicine, Norwegian School of Sports Sciences, Oslo, Norway

Correspondence to
Dr Stephen W West;
S.West@bath.ac.uk

The expansion of technology in sports has dramatically altered the experiences of fans, practitioners and athletes. Innovations such as cloud computing, wireless connectivity, wearable sensors and artificial intelligence (AI) have expanded the scope of athlete monitoring on and beyond the field of play. Measuring more variables on more athletes and in more sports than ever has created a data boom. Big data are not, however, simply a consequence of technological advancement; it has wider consequences both within and beyond sports.[1] While these changes have provided new opportunities to impact player welfare and performance, they require proactive data management review and reform, particularly in sports science and medicine. The era of big data in sports is upon us, and it is a game changer.

The 'Three Vs' of big data capture in sport (ie, variety, volume and velocity)[2], combined with significant multisector investment, will result in unprecedented and paradigm-shifting access to athlete data. Competitive advantages will not be found in capturing data that competitors cannot, but in the ability to manage and transform data into actionable insights. Many sports science and medicine professionals are *de facto* athlete safety, health and data stewards. Therefore, these fields must re-examine the questions asked, approaches used, and outcomes prioritised to unlock and maximise the potential of the modern big data era in a way that is grounded in athletes' rights.

While athlete monitoring technologies have increased across all dimensions of access,[3] access is not automatically distributed fairly or justly across sports stakeholders. Data ethics, governance, policies, procedures and best practices must be developed with rights-based and justice-based lenses to mitigate the perils of big data. It is argued that these hazards have already begun to impact athletes across the global sports landscape.[1]

Data are captured, analysed and interpreted by many individuals, organisations and institutions across the socioecological spectrum (figure 1). The lines between data rights, ownership, management, access and consumption are often blurred. With so many stakeholders in the pipeline, responsibility for data ethics, governance, policy and practice standards can be ambiguous. Examples include third-party vendors being granted access to data under specific terms of use, or the vulnerabilities that can arise when employees leave or join an organisation, affecting how data are accessed and managed—the dynamics of sports present additional challenges in personnel mobility. Athletes and staff can move between organisations (eg, two-way players, players on assignment) or exit the sport entirely (eg, graduation, waived/fired, retirement, etc). We must ask questions about the policies, procedures and practices for securely managing, sharing or deleting data within fluid and often unpredictable contexts.

Data governance legislation varies widely around the world. In the USA, for example, personal health information privacy regulation is covered in the Health Insurance Portability and Accountability Act (HIPAA), while in Europe, the European Union General Data Protection Regulation (GDPR) applies, which covers 'any information relating to an identified or identifiable natural person'.[4] The GDPR was enacted in 2018 and prompted a sea change in data protection legislation and regulation worldwide, including in South America, Africa and Asia.[5 6] As high-performance and professional sports continue to globalise and technologise, the data practices across sports operations must be internationally compliant while identifying and mitigating collective and individual athlete health and welfare risks. An important concept within these data legislations is that of 'Data Minimisation' (ie, '*collecting the minimum amount of personal data that you need to deliver an individual element of your service*'[7]). Given the recent tech boom, sport has been described as suffering from a 'collect it all' ideology[8] whereby data are collected because we can, not because we need to. Often, this is done to inform what may be possible in the future (eg, with advances in AI). This could

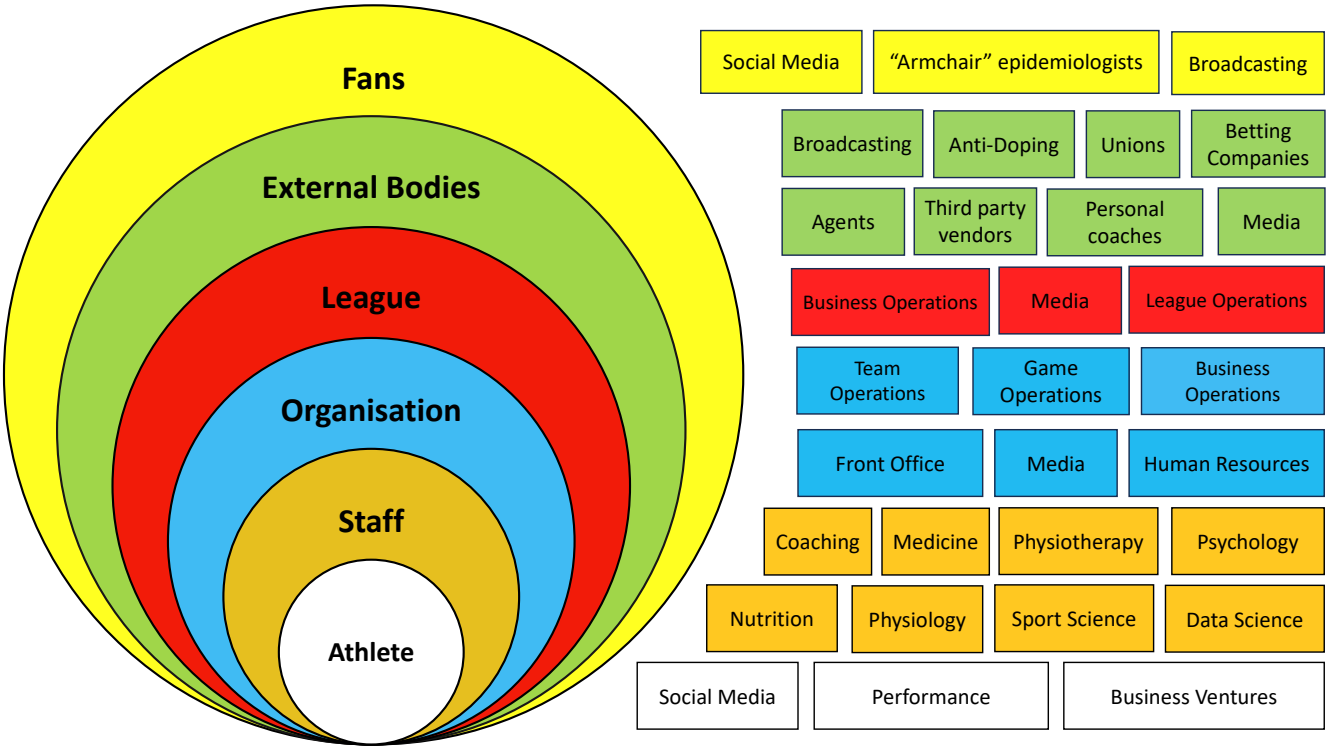

Figure 1 Data contributors, conductors and consumers.

be considered appropriate by those who set to profit from such future work; however, it could be seen as problematic by those who are at risk of being manipulated, disenfranchised, disempowered or exploited by the data.

In many cases, this 'collect it all' ideology is cloaked under the veil of load management. Still, practitioners should carefully consider what is driving these collection processes and whether the data captured are necessary, useful and actionable. Further, when dealing with potentially sensitive information, it is critical to distinguish between data based on evidence and not, particularly regarding informed consent. For example, with menstrual cycle tracking, athletes should be provided with greater information on why the data are collected and what efforts have been made to achieve data minimisation.[9] In considering these important reflections on current data practices, anecdotal evidence would suggest that most organisations do not follow the principle of data minimisation.

In the daunting world of big data, new technology, and AI, it is our responsibility to protect the rights of the athletes we support and ourselves. We can achieve this through improved data protection literacy and critically reviewing, reforming and, where necessary, reconstructing the data capture processes and infrastructures we use. Practitioners should be cognisant of their role in this space and the wider data economy, critically reviewing and reflecting on how its evolution impacts our ability to provide and receive informed consent.

Like many industries, sports is under pressure to 'keep up' with the world of big data, AI, and technology. In a world where clicking on 'I agree' has become nearly automatic, many of us have become desensitised to the evolving risks associated with the access we authorise. While we should embrace innovation, athlete welfare and protection should be paramount. Transparent and responsive practices related to data operations must be at the heart of our conversation. Data protection and governance are not solely a sports issue. As an early adopter of technology, sport must ensure its approach is rights based, athlete centred and value aligned, with standards of operation that are as high for the processes off the field of play as they are for the performances on it.

X Stephen W West @westy1609, Jo Clubb @JoClubbSportSci, Tracy A Blake @tracyablake, James Fern @JamesFern83 and Torstein Dalen-Lorentsen @torsteindalen

Contributors SWW and TD-L were responsible for the concept of the article. All authors were engaged in the production, editing, review and completion of the article.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient consent for publication Not applicable.

Provenance and peer review Commissioned; internally peer reviewed.

ORCID iDs
Stephen W West http://orcid.org/0000-0001-5800-7000

Jo Clubb http://orcid.org/0000-0002-6509-7531
Tracy A Blake http://orcid.org/0000-0002-1888-2940
James Fern http://orcid.org/0009-0007-8708-3265
Harry Bowles http://orcid.org/0000-0002-0541-3926
Torstein Dalen-Lorentsen http://orcid.org/0000-0003-4062-7601

REFERENCES

1 Bowles H, McGee D. Sport, surveillance and the data economy: an expanding horizon for research and governance. *International Journal of Sport Policy and Politics* 2024;ahead of printdoi:1–9.
2 Laney D. 3d data management: controlling data volume, velocity, and variety. *Application Delivery Strategies* 2001.
3 Penchansky R, Thomas JW. The concept of access: definition and relationship to consumer satisfaction. *Med Care* 1981;19:127–40.
4 Union E. What is considered personal data under the EU GDPR. Available: https://gdpr.eu/eu-gdpr-personal-data/ [Accessed 6 Mar 2024].
5 Carrillo AJ, Jackson M. A comparative law study of the EU's general data protection regulation's impact in Latin America. *ICL Journal* 2022;16:177–262.
6 Lovells H. Recent developments in African data protection laws - outlook for 2023. Available: https://www.lexology.com/library/detail.aspx?g=baef72ee-10bd-4eb9-a614-a990c236bb45#:~:text=To%20date%2C%20thirty%20six%20out,it%2C%20the%20latest%20being%20Niger [Accessed 6 Mar 2024].
7 Office ICs. Data Minimisation. Available: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/age-appropriate-design-a-code-of-practice-for-online-services/8-data-minimisation/#:~:text=Data%20minimisation%20means%20collecting%20the,child%20actually%20wants%20to%20use [Accessed 6 Mar 2024].
8 Powles J, Walsh T, Alderson J, *et al*. Getting ahead of the game: athlete data in professional sport. 2022. Available: https://www.science.org.au/files/userfiles/support/reports-and-plans/2022/getting-ahead-of-the-game-athlete-data-in-professional-sport.pdf
9 Moore IS, Crossley KM, Bo K, *et al*. Female athlete health domains: A supplement to the International Olympic committee consensus statement on methods for recording and reporting Epidemiological data on injury and illness in sport. *Br J Sports Med* 2023;57:1164–74.