

Module 2. Introduction to Metric Data Analysis

Unit 2.1 Analytics: The Alamar approach

The use of the term analytics has increased in the sports world in the last decade. As Jayal, McRobert and Oatley indicate (2018), this is not another name for what was previously known as **analysis**, but it is linked to decision-making among athletes, coaches and managers.

Benjamin Alamar (2013) draws conclusions based on a survey of the use of analytics in sports institutions ("SAUS" Survey) in terms of its impact on decision-making processes.

Figure 1: Benjamin Alamar



Source: [Untitled image of Benjamin Alamar]. (undated). Retrieved from: https://www.linkedin.com/in/benalamar?trk=public_profile_browsemap_mini-profile_title.

This SAUS survey was conducted in NFL (American football), MLB (baseball), NBA (basketball) and Premier League (English soccer) clubs. Among the main findings, Alamar indicates that there are often different visions within the same club regarding their perception of the level of quality provided by the analytics. It is common for two executives of the same club to have different views on the level of adaptation and usage of a database or the application used by the club. In other words, in professional clubs, there is often no internally shared vision of the competitive advantage in terms of the data obtained, in comparison with their rivals.

The movie "Moneyball," based on the methodology employed in baseball since the 2000's and used by the Oakland A's, was met with less resistance from the public, press and managers regarding the use of data in sports. Paul DePodesta was manager Billy Beane's chief data assistant with the A's.

Figure 2: Paul DePodesta



Source: [Untitled image of Paul DePodesta]. (2017). Retrieved from: [https://commons.wikimedia.org/wiki/File:Paul\(36380255290\).jpg](https://commons.wikimedia.org/wiki/File:Paul(36380255290).jpg).

An example of the possible transfer of analytics to other sports is Paul DePodesta's transition from baseball to American football, in which he is currently the director of strategy for the Cleveland Browns in the NFL.

The American approach to sports data has been more industrious than the European one, which has focused on research, understanding of the game and a long-term perspective, despite authors such as Peter O'Donoghue (professor at Cardiff Metropolitan University), who linked his studies to decision-making. European high-performance and training centers typically have research laboratories that feature more established and solid knowledge, but which face greater difficulty in terms of their direct and immediate relevance in the player acquisition market or game plans.

2.1.1 Whitepaper: Davenport's vision

Thomas Davenport, an analytics expert from the corporate world, identified sports as the most promising new field for the application of data. He carried out a project featuring 25 interviews with experts from both European and American professional sports and soccer teams, including not only coaches and executives of these clubs, but also information systems experts.

One of the main factors for success identified by Davenport (2014a) was quick and direct access to video. Not having to rely on exchanging videos with other teams, thanks to the availability of corporate league platforms, can be a pillar on which to promote a culture of data, both for the *customized* registry and indexed visualization by means of metadata.

In his conclusions, Davenport (2014a) divides data-based decision-making into 3 different areas:

- **Sports performance**

- Identifies the following key differential factors: the use of positional data (tracking), the involvement of players in the visualization of reports, or the creation of proprietary databases with *customized* variables.
- **Sports management**
 - Primarily focused on the study of the fans, or those who attend the stadium or sports facility, by means of Wifi networks that allow for the study of their movements and consumption, as well as the classification of ticket prices or marketing activities.
- **Health**
 - Classifying injuries and being able to predict them is a priority.

Some of the factors that Davenport (2014a) identifies as good practices in terms of metric analysis are as follows:

- Alignment and common vision among decision makers.
- The leagues are responsible for coordinating investments.
- Clubs buy and acquire data that then belongs to them.
- The clubs search for the advisory services of consultants, both external and within the industry; establishing partnerships.
- Establish projects that allow volunteers to participate in the club, offering alternative perspectives in data processing.
- Create combined indices that unite sports and financial variables.
- Explore the possibilities of predictive and prescriptive analysis.

2.1.2 Game analysis: the adaptation of O'Donoghue

Peter O'Donoghue conducted a review (2014) of different available software solutions for a play-by-play analysis that allows for building on the knowledge of the game, such as FocusX2, SportsCode, Dartfish, Excel, SPSS or MATLAB. The author also revealed the possibilities of positional soccer data with X, Y coordinates of players and the ball, provided by Prozone (currently known as STATS). His adaptation of game analysis can be studied in the periodicals of the academic journal, the International Journal of Performance Analysis in Sports, of which he is the editor. O'Donoghue (2014) describes scenarios in which data recording is done manually by an observer, as well as those in which data is imported from data providers.

Lebed (2017) reveals the progressive transition that takes place when transforming data into knowledge, which has an impact on the subsequent step; that of wisdom, which is necessary for decision-making in sports.

Annex I

Case study on the introduction to metric data analysis.

In June 2015, Bloomberg Businessweek magazine published a monograph written by author Paul Ford (2015), entitled "What is code?," exploring computer programming languages, coding and development. The issue continuously alternated between two blocks of content: one on information, and another on fiction concerning the potential application of said information in the software industry. We used this idea to recreate a possible, practical and fictitious scenario of the sports industry and metric data analysis. Antonio waits his turn in the room for his job interview. In 2011, tablets were still slow devices that weren't widely used, so he has prepared some notes on his BlackBerry, which still competes with Android in the market. He has printed some charts and tables that he carries in a folder with plastic dividers, in case he has an opportunity to present them in the interview.

The BlackBerry agenda sends several notifications that inform him that the interviewers will arrive late. Finally, it's his turn. It is a company that manages the then in force "Data Protection Act," of a club of elite indoor sports players, which does not have a soccer section. He contacted them by email and spontaneously called a few weeks ago to propose a research project in data management, related to the doctoral thesis that he has recently begun at the university in the same city as the club.

After five minutes of conversation, the decision is made and communicated to him: Antonio will start working for the data company in the club the following day. His role and salary are not made clear, but they want to have him there "to do something with the data". He will collaborate with the reserve team in recording physical tests, injuries, etc., although he has not yet mentioned that he is passionate about competition analysis.

It is the first training session, and Antonio is the first to arrive, wearing an office suit in the training area, where there are coaches and players wearing shirts and shorts. He is asked to sit in a chair full of resin that destroys his suit, and after introducing himself, the company representative of the club instructs him to manually record players' attendance in the training session on a photocopied sheet.

On the second day, he receives sports clothes from the club with the logo and his initials screen printed (he made a good impression on the reserve team, according to conversations with the coaching staff in the previous training session). He has an open Excel in the BlackBerry that he sends by e-mail to the assistant coach of the team at the end of the training, detailing the attendance and absences, which gradually begins to replace the photocopy he received.

There is still much to do, but he has not received any sort of mission, only that something must be done with the data; generate or store it, and draw some conclusions about the findings. His daily life becomes further detached from the objectives of his doctoral thesis, which he finally abandons (or rather, does not initiate) in order to focus on the club.

He recalls the name of the American sports magazine "Insight". That's what the club and the company want from him. To generate insight; a profound vision that has an impact. The coaching staff of the first team begin to take an interest in the individual training and game notes that he records for the reserve team, and they start asking him to do the same for them. A few days before the end of the preseason, he receives an invitation from the physical trainer of the first team to collaborate with him in conducting a test.

One of the players is late for the test training, and the coaching staff begin to smile when they see Antonio record a "semi-foul" for attendance on his Blackberry. The reserve team training session begins in the adjacent track, and Antonio is greeted with a good-natured taunt, "you don't want anything to do with us anymore," by the coaching staff of the reserve team, to which the first team staff respond, "we have promoted him just like the other guy last summer, who started the preseason with your team".

After tabulating the data and presenting it to the physical trainer, Antonio is instructed to create training groups based on the observed endurance values. His idea gets approved: in addition to creating 3 groups that are formed by the trainer, he uses a traffic light color code, in which green indicates that they are above the 75th percentile, according to the scientific bibliography in that sport, while red indicates performance below the 25th percentile. The presentation is printed on paper after each test, until 5 months later, mid-season, he creates an evaluation in which he has 5 minutes to show the coaching staff, directors and managers some of the results of his work. This is the first time he has been asked to create a PowerPoint presentation.

The old club projector does not work well and cannot discriminate between reds and greens, instead giving them an ochre color that makes it difficult to understand the data presented. However, the coaching staff did not expect such candid assessment in a meeting in which they feel that their work is being negatively judged by the endurance data presented, among other aspects, by both the tests and tracking. Antonio did not previously consult the physical trainer regarding whether or not he should show the data from that test, and regrets not having asked for permission.

Since that afternoon, Antonio has lost the trust of the coaching staff, but during the summer the directors offer him a position in the club, and he joins at the beginning of preseason. Once again, he does not receive any specific objectives; his work seems to be that of an "on-call" data analyst who must generate reports for any manager who requests it.

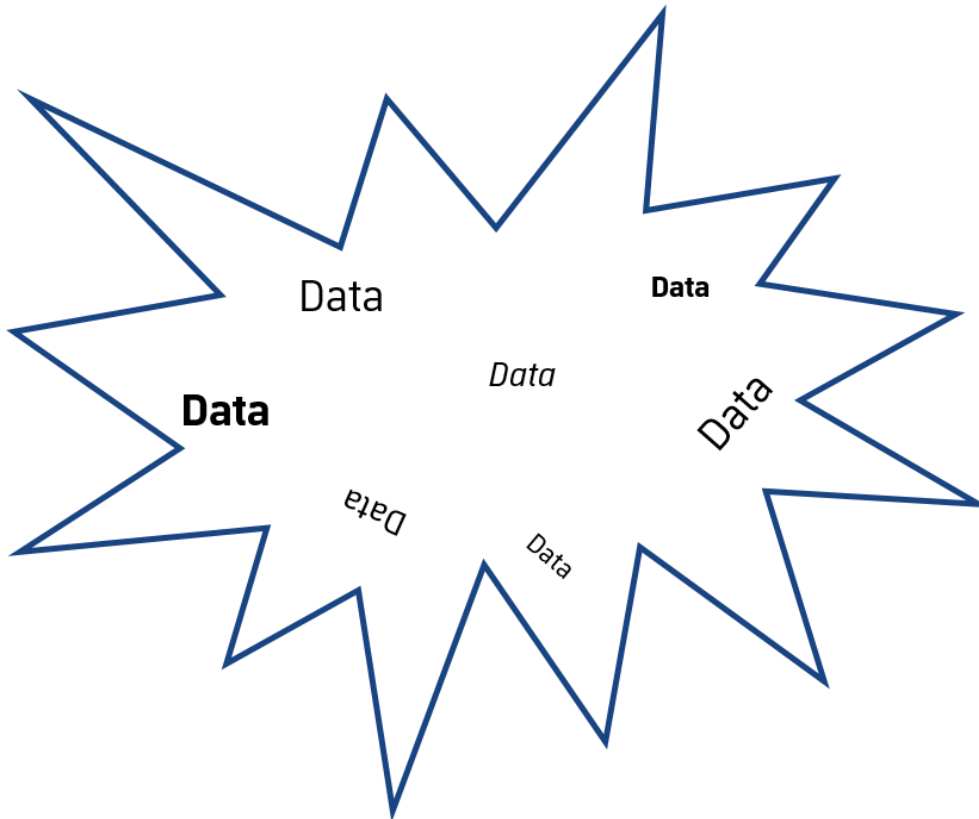
A week after joining the club, Antonio has received requests for reports from the financial department, the presidency, management, medical department, etc. on sports data that can be linked to player contracts, in order to calculate their efficiency or to determine the appropriate changes in management processes. However, he has not been contacted by anyone in charge of the sports department, despite working with data. He has broken one of the "unwritten laws" by airing their dirty laundry, even if it is in the form of test-related data.

Unit 2.2 Databases

2.2.1 Storage: data lake

Once the sports data has been collected or recorded, the next step is to establish a single repository in which they are digitally stored. This concept is known as a data lake.

Figure 3: Data lake



Source: Prepared by the authors.

The data lake can be set up on the club's own server or through cloud access storage, with advantages and disadvantages in each case.

If he opts for a physical data lake, it will be difficult to access important videos and files when a game is played in another city or when members of the coaching staff try to access it from home, although it is possible (but slower) with solutions such as a VPN (virtual private network).

If the chosen solution is a data lake cloud, the maintenance and hosting costs are lower for the same volume of data, as well as enabling access from any location. As a drawback, the speed of data access (which is particularly important when accessing files with several

Gb of data, for example, videos or Raw Data files from games or training sessions) gradually increases the price of the cloud.

In any case, the key objective is to avoid data silos in which club information is stored in different repositories that are not connected to one another, therefore the data lake is a suitable solution for this essential function.

2.2.2 Structuring: data warehouse

Once the data has been centralized and stored in basic formats in a data lake, it is necessary to carry out structuring and hierarchy by means of a data warehouse.

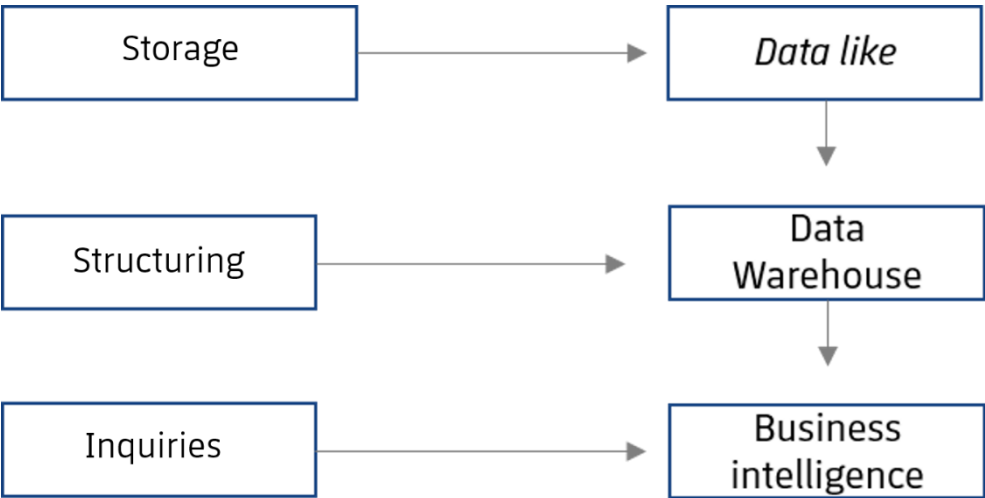
The data lake will enable ETL functions (E=Extraction, T=Transformation, L=Load) of data, organized in subunits known as data marts.

In order to carry out the transformation functions in the data lake, a preliminary task of defining and classifying basic data is needed in order to enable the appropriate governance structure.

2.2.3 Inquiries: Business Intelligence

After its storage in the data lake, along with the preparation and hierarchical organization in the data warehouse, the "data consumption is employed" through consultation using a Business Intelligence solution, with applications that provide visualization and filter possibilities without the need for advanced technological knowledge.

Figure 4: Database Usage Process



Source: Prepared by the authors.



Annex II

Database case study

It is the beginning of Antonio's third season with the club, and the following mantra can be heard throughout hallways and club offices: "ERP, ERP, ERP." There is a long-standing concern among the club directors about the loss of data when an executive or technical staff member leaves the club, taking with him his computer and sensitive data in spreadsheets and non-centralized applications. The need for a **centralized database** was first discussed a few months ago. In the initial meetings to address this objective, the finance department explained to the sports department that they can make use of the available solution because it is "modular". Inquiries made to IT suppliers about how long it would take to provide a record of training exercises were initially met with silence, casting doubt on the suitability of these solutions.

Antonio was commissioned to carry out market research. He presented several solutions with a SWOT analysis and price estimates in a subsequent meeting, after which the club ended up leaning towards the solution proposed by IT (he clarified that he is not an IT specialist, but rather an administrator with IT knowledge who works in that department of the club). Such solution was previously evaluated by Antonio, in which he indicated that on a positive note, they would be able to obtain quick wins. On the other hand, the chosen solution would also quickly hit its limit, as it does not currently have imports or exports, or even the possibility of communicating with other databases of the club. Its main weakness is that the goal of centralizing all the data generated in the club would not be met.

A few weeks later, Antonio received three pieces of good news:

- Internship agreements with universities were approved that will allow for the participation of students from different areas of study who require ETL data training, who will also assist in obtaining reports on injuries and physical performance.
- He received approval to hire an assistant to help him with his daily work.
- Next season, the Federation will supply a new computer application that will provide a wide range of data (in real time and post-game) all about the competition, which they do not currently have.

The pace of work has dramatically increased in the offices. Student interns, dressed in clothes temporarily loaned to them during their two-month stay, walk around the training facilities with corporate tablets, recording the subjective perception of the players' effort and storing this information in the newly purchased application. The new assistant is in charge of the routine coordination in the Technology Center, since Antonio must

frequently visit the capital, where the league has its headquarters, for thematic and sectoral meetings in which he represents the club.

A meeting of the Board of Directors, which takes place on one of the nights upon returning from the meetings in the capital, unexpectedly results in a change-up at the top level of the Board of Directors; producing an almost complete redistribution of the management and executive bodies in the new structure of the club.

Faced with this new situation, Antonio presents his briefing at the monthly management meeting. He enthusiastically presents the two new computer applications that the Federation has created, which will give the club a competitive advantage, thanks to the work previously carried out.

The club's culture has changed: Antonio reports that at least two new devices must be purchased to run the new applications; an iPad for live game software and a powerful Windows computer for post-game purposes. The club does not currently have either device, and applications must be paid for by all clubs belonging to the highest competition level of the Federation.

The season begins and neither the iPad nor the "powerful" Windows computer has arrived. He visited the manager's office several times the previous week, where he was told that they are studying the costs. Antonio has borrowed an iPad from a doctor for the first league game to install the live application.

On Monday, Antonio receives a call from the Federation asking him why he hasn't yet opened the key associated with the IP that activates the new post-game application. He explains that the club has not yet equipped its IT department with the necessary requirements. The Federation tells him that his is not the only case.

After a few months, the computer is delivered, and it is already round 16.



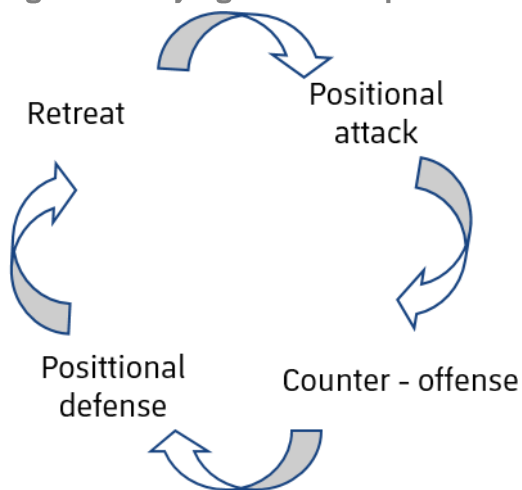
Unit 2.3 Data selection criteria

2.3.1 Whitehead: to know how to observe is to know how to select

An appropriate selection of variables to be observed is usually determined by external data providers, who use as wide a selection as possible to incorporate the various perspectives and needs of different clients, in which a structuring of the playing models and stages is often missing.

A starting point from which to select the metrics to include in an analysis model is the establishment of the stages of the sports model in question. For example, in team sports, the 4-stage model is typically used.

Figure 5: Playing model in sports



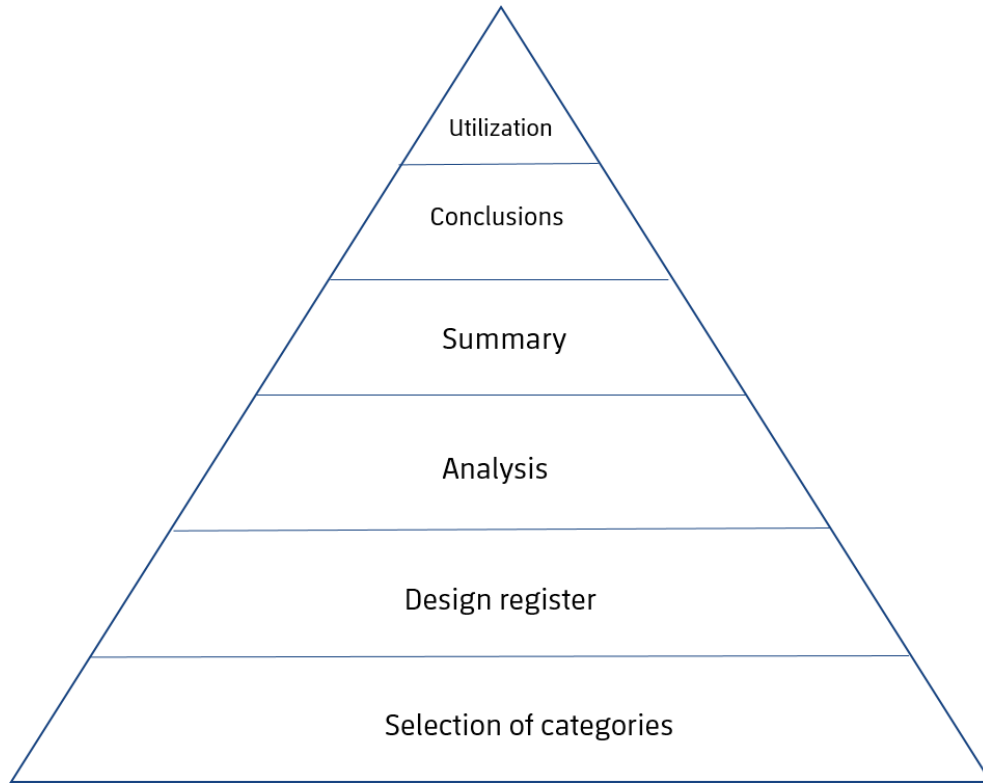
Source: Prepared by the authors.

Whitehead (1967) indicates that "to know how to observe is to know how to select" (p. 28). All models are approaches that omit the less relevant aspects, but allow for an understanding of a phenomenon. There are two fundamental approaches in selecting the variables to be observed:

- Quantitative, in terms of the impact of the variable on the final result.
- Qualitative, referring to the key relationships in the chain of actions, although the examined variable itself does not have a high impact on the final outcome.

After selecting the categories based on a suitable playing model, the following are required:

Figure 6: Database utilization pyramid



Source: Prepared by the authors.

- **Design register**
 - A live game registration app that provides information during the break does not share similarities with a post-game analysis app, which features more available operators and time. There are personnel and time constraints that can even affect the previous step of category selection.
- **Analysis**
 - Once the registration tool is adjusted to the selected categories and made available, the observation itself begins with the appropriate digitalization.
- **Summary**
 - The ability to openly filter and readjust any type of query is not possible in many apps and software solutions that *preset* the type of accessible report. Being able to access any type of filter and combination of registered variables gives meaning to the effort previously made in the analysis and registration.
- **Conclusions**
 - The transformation of data into knowledge leads to descriptive rather than prescriptive findings from reports.

- Utilization
 - The impact of metric data analysis on training or competition can be seen, based on the coaching staff's interpretation of previous findings.

2.3.2 The 5 types of sports data

Any type of data available in the sports industry can be classified within the following 5 blocks:

- *Eventing.*
- *Tracking*
- Biological.
- Financial.
- Video.

The eventing data indicates sequences of technical-tactical actions that can be recorded, such as kicks, passes or dribbles in soccer, within a given time frame. It also includes data, such as the date and place of the sporting event or the names of athletes.

As for tracking or positional data, these are spatial coordinates occupied by athletes, players, instruments or objects in the sports facility at certain frequencies (a sampling frequency of 1 Hz would indicate one data capture per second, whereas 50 Hz would indicate a capture every 0.02 seconds).

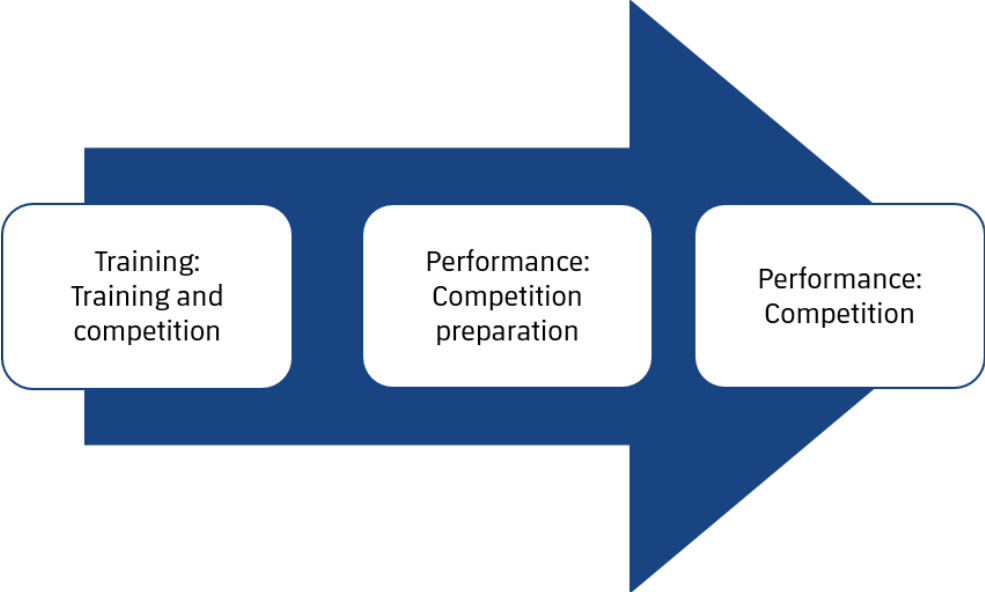
Biological data refers to all biomechanical or physiological analyses (blood, weight and body composition measurements, injuries, heart rate, strength output) of athletes. This type of data is also known as vitals in Central Europe.

The financial data, most often used within the American sports industry, deals with economic amounts referring to salaries, signings, transfers, etc.

Video being considered data is a change of approach that has occurred with its digitalization, but which has not yet gained traction among the habits of most European sports institutions, which can provide a competitive advantage for those clubs that implement it. Linking the video to a filter of variables in any database to efficiently transmit information to players is often overlooked.

There is another cross-cutting classification of data that influences the type of findings and uses. Hohman, Lames and Letzelter (2005) see training as a **preparation for competition**. If we expand on this approach, competition during the sports training period for young players is a different type of competition than high performance, in which other perspectives should be applied when utilizing the data.

Figure 7: Cross-cutting approach



Source: Prepared by the authors.

2.2.3 Standardization and *customization*

The standardization of the 5 types of data referred to above has traditionally been carried out by private companies that were the first to provide a niche market, or those that achieved leadership in the area of competition; resulting in imitation or copying by other stakeholders in the sports industry.

Table 1: Examples of the 5 types of sports data

Examples: sports data standards	
<i>Eventing</i>	F24 (Opta)
<i>Tracking</i>	EPTS (FIFA)
Biological	OSICS
Financial	<i>Financial Fair Play</i> (UEFA)
Videos	PAL, DCI 4k, H.264/MPEG-4 AVC

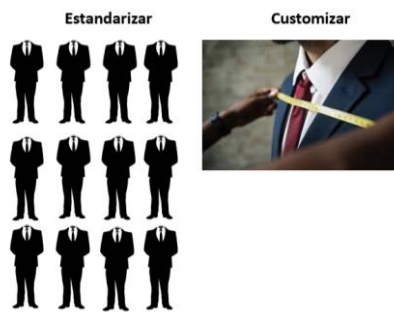
Source: Prepared by the authors.

In recent years, institutions such as FIFA and UEFA have provided standards that may indicate a new path forward in sports data leadership.

In any case, *customization* is necessary in order to gain a competitive advantage in sports.



Figure 8: Graphical difference between standardization and customization



Source: Prepared by the authors.

The key aspect lies with the starting point of *customization*: starting from scratch can be inefficient, as an admissible performance can often be obtained by creating new combined indices from isolated metrics in the standards.

Annex III

Case study of data selection criteria

After a long period of uncertainty, the I+D+I (Investigation + Development + Innovation) department of the club can analyze the tracking and the eventing data from the games, using the app that the Federation has provided. Interns carry out exports every Monday in csv format of the play-by-play data, which will subsequently be processed in Excel in order to generate insights not included in the standard Federation reports. The possibilities of combining data are almost unlimited, there are sufficient resources, and communication with the new coaching staff is fluid. The only potential limiting factor is time or efficient time management.

The second coach of the first team usually asks in the hallways, "Is it possible to calculate...?...Do you know...?", and thus a glossary of queries is created and incorporated into a "Manual of Metrics processes of the first team," which attempts to systematize the most common stream of inquiries, although there are always unforeseen questions hours before a game. Rankings, reference values, individual player contributions, etc. are systematized into post-game reports, and the pre-game reports are made by taking the opponents into consideration.

The medical department begins to utilize accelerometer data that allows for bimonthly meetings on training loads and games. It is an opportune moment for the data and its applicability, as the team is favorably ranked, which allows them to play in the European competition at the end of the season. However, the main sponsor of the club goes bankrupt and the future is now uncertain. Vacation involves a series of inquiries to bank apps to check if money has been received from any of the overdue paychecks, with rumors of layoffs and calls asking if he is willing to accept a 30% decrease in salary. Finally, after

the second day back at the club after his vacation, Antonio and five of his colleagues meet in the office of the managing director, who tells them that they are being let go, and sends them to the administration department to sign out upon leaving.

It turned out to be good timing: one of the eventing and tracking providers had been asking him to join his R&D Department for some time, although the job would also involve sales to other clubs. After visiting the European headquarters, Antonio accepts. He will be responsible for the Iberian Peninsula with one of the largest providers of sports data.

The company already has good data standards; in fact, it has created a format that has been replicated and copied by their competitors, who operate with lower costs, and are losing a market share which nearly reached 100%. Antonio must carry out interviews in the clubs of the "Iberia Zone" that allow him to *customize* reports, as requested by the coaching staff. His meetings are productive and allow him to understand the mentality of rivals that he has studied and faced for five years. However, when he makes a call or sends an email with the price of the new product, after a "demo" of several weeks, none of the clubs purchase it.

It has been a few months of learning more systematic processes within the company, of learning English and Portuguese, along with the playing models of the different coaching staff. The company has been patient about making the first "sale," but after the first semester, it is taken over by a group that has also acquired their main competitor. In a few weeks, Antonio must study the standardized methods of the former competitors, which replaces the *customized* strategy that was previously applied.

A Federation from another continent contacts him during an eventing and tracking conference. They offer him the opportunity to join their new analysis department, with the aim of implementing data analysis processes "from scratch," in a country that will barely qualify for a World Cup, let alone the Olympic Games. They allow him to work remotely most of the time and travel occasionally every two months. He decides to change jobs.

The Federation that Antonio works for now has provided him with basic resources to carry out his new "Project for the implementation of data-driven decision-making," inspired by the "data-driven organizations" that he has come across in conferences and symposiums. The Excel sheets managed by the Federation become blocked when they reach a size of 7-8 megabytes, when entering names and dates of birth of all of the registered players. It is therefore necessary to have an SQL database as a solution, created by the Federation itself, for which a developer is hired and who becomes Antonio's main ally.



Unit 2.4 Use and transfer of data

2.4.1 Big Data or Smart Data?

As Davenport (2014b) indicates, the term Big Data has suffered from overuse in the field of sports, often due to commercial pressure. We can classify Big Data as the field of data in which **the three V's** converge: volume, velocity and variety. A fourth V has recently been added: veracity. However, there are no agreed-upon limits with which to classify information as Big Data. Big Data is also often associated with unstructured data, or that which does not follow a conventional structure of rows and columns.

To a greater extent than Big Data, it is worth having access to Smart Data in sports, as these techniques enable access to and use of pertinent data.

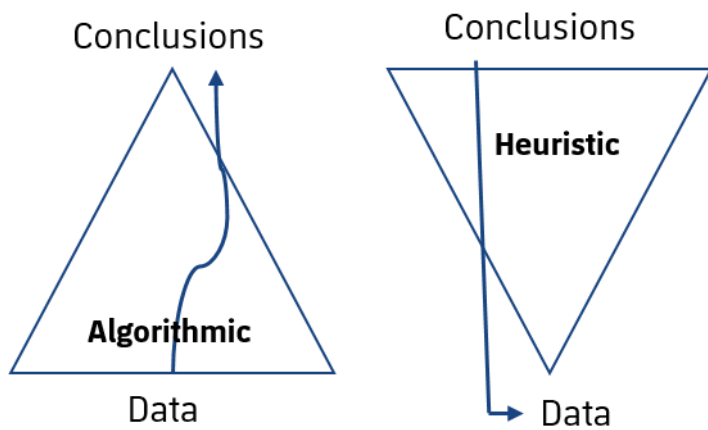
Certain terms are often used that can be distinguished from the former:

- Artificial intelligence: encompasses methods that enable learning from a data structure to achieve an objective.
- Machine learning: a branch of artificial intelligence that detects patterns without prior instructions.
- Deep learning: another variant of artificial intelligence that assimilates data representation.
- Neural networks: applied to problem-solving, it allows for integrating data representation.

The most common approach taken by coaches and coaching staff is heuristic-inductive, which is based on a number of insights and conceptions of sports competition that can be compared with data.

On the contrary, data managers of the clubs tend to approach the sports phenomenon in an algorithmic-deductive way, in which one of their priorities is to avoid bias and transfer data associations as findings only when they are overwhelmingly confirmed.

Figure 9: Different approaches to data processing



Source: Prepared by the authors.

2.4.2. Know-how in sports

The process of maturity and improvement of a coach usually begins with an initial period in which there is no discrimination between what is important and what is urgent, nor what is most relevant and what is secondary. The sooner the most relevant factors associated with sports performance are discerned and strengthened, the better the adaption to the training method and preparation for competition.

Along these lines, a database or dashboard that influences performance must particularly take into account the most relevant game scenarios, which do not often occur.

Table 2: Example of a dashboard linked to sports performance

Relevance frequency	Little Relevance	Very Relevant
Infrequent	<p>Infrequent and irrelevant</p> <p>Ground ball, drop ball in soccer.</p>	<p>Infrequent and Very Relevant</p> <p>Shots on goal in soccer.</p>
Very Frequent	<p>Very frequent and irrelevant</p> <p>Passes between centers without pressure in soccer.</p>	<p>Very frequent and very relevant</p> <p>Passes in the last 1/3 of the field in soccer.</p>

Source: Prepared by the authors.

It is a question of systematizing what is relevant.

In the general scope of business, Charan (2007) identifies eight criteria of "know-how" that can be adapted to sports:

- Identify the most relevant aspect(s) of sports competition.
- Detect threat(s) from the competition.
- Lead athletes in a sociable way.
- Evaluate players and coaching staff.
- Achieve cohesion through team-building activities.
- Establish objectives.
- Establish priorities.
- Anticipate the surroundings and social pressure.

2.4.3 Competitive advantage

Two reports with the same findings that reveal the same data may differ in presentation and visualization, and apart from the aesthetic aspects, one will successfully stand out to the coaching staff while the other will not. With this in mind, it is worth noting the different aspects of visualization categorized by Jayal et al. (2018) in sports:

- Interactive.
- Simulation.
- Modeling.
- Geo-spatial.
- Spatial-temporal.

What will be the next competitive advantage detected that will lead to victories and success? There are specific cases that have had an impact on sports history, such as the increase in pace of play in handball.

We can identify success stories over the course of the past decades in soccer:

- 1920's: professionalism.
- 1930's: schools and playing systems.
- 1950's: talent in decision-making.
- 1970's: universality of posts.
- 2000's: transitions.
- 2010's: generation of numerical or temporal superiorities.

Most of the time, it has been the intuition of players or coaches that has paved the way towards innovations, which have signaled an era of competitive advantage. On the other hand, there are new strategies based on metric data analysis, such as the recent increase

in 3-point shot attempts, which was the idea of Morey, manager of the Houston Rockets (NBA) since 2007.

Annex IV

Case study on data usage and transfer

The developer who was hired uses Python programming to transform the tracking files received at the Pan-American Championship, and responds to queries made by the coaching staff through graphic representation of the distances of the defensive players as the game progresses. The eventing files are processed with R programming and SQL solutions, which allow for observing the reasons behind a favorable or unfavorable lead at a given time in the game. However, feeds (original standardized data files) are not available for "home" competitions or continental youth tournaments, for which an Android application is used that allows the coaching staff to register basic eventing data of the games themselves.

During one of Antonio's face-to-face visits, the process of creating this app serves as an excuse to learn of the know-how of the Federation's technical director and coaches. He thinks that the level of inquiries regarding the defensive stage of the game is far higher than that of Europe, so he must study the game even more in-depth in order to keep up with the app's design. However, the level of inquiry is very basic in the offensive stages (including the counter-offense), and after consulting the benchmarking of the teams from this continent in world competitions, he believes that it can be a key element for a competitive advantage in the next Olympic cycle.

After a few months, Antonio returns after crossing the Atlantic once again. After presenting the Android app, he takes advantage of a moment of quorum to present the rankings of the offense, defense, retreat and counter-offense of the teams in the last Pan-American and World Championships. There is a clear trend in the Power BI, QlikView and Tableau dashboards he has prepared: a quick pace of play has provided an advantage in most of the games the team has played in the last 5 tournaments. The coaching staff have their reservations about this, but feel that it is an insight that enables them to increase the performance and pace of play of their teams, which can benefit from the type of player in that country.

Both men's and women's teams increase their playing speed in subsequent tournaments, while the men's teams experience an increase in loss of possession and a decrease in performance, and the women's teams surpass expectations. Players start to transfer over to European leagues, which, once consolidated, strengthens and refines them; enabling an even greater increase in the pace of play with better precision. They qualify for the

Women's World Cup, while the men's teams return to a slower pace of play and slightly improve their performance.

During one of the reporting meetings of the season, findings were presented on why increasing the pace of play has been beneficial for the women's teams, but not for the men's teams. Hypotheses are presented regarding the lack of male players in European leagues, while there are more female players in European leagues than in domestic ones. At the same time, slight discrepancies have been detected in the refereeing criteria for women's and men's competitions.

After taking note of the hypotheses proposed by the trainers, Antonio performs a regression analysis through the SPSS program of the variables available in Pan-American championships. However, there are too few games to draw conclusions, so he calculates the size of the sample needed to create a presentable report for the Federation.

In his calculation, Antonio concludes that there are few competitive games against rivals that would allow the teams to improve in the Olympic cycles. He recommends increasing the number of these games, for which a greater investment would be required, especially in the men's teams, where the majority of players do not compete in European leagues.

There is no available budget to carry out this objective, but the Federation decides to establish agreements with European second division teams, which would lead to a greater number of players in these leagues. Antonio estimates the potential improvement in performance by having at least 5 players in these competitions.

After the agreement has been implemented, players begin to arrive in the European leagues, which are initially grouped into two neighboring countries. Antonio is responsible for the individual tracking of these players, and is aware that for most of his career he has devoted his time to collective and team data, but not individual data. He should learn and make the most of the know-how of the Federation's coaching staff regarding individual aspects of the game, and design new registration tools. He should also implement a "market calculator," which estimates the price a player can fetch after his/her time spent in Europe; both for a potential transfer in which the Federation obtains training rights, and for the potential salary that clubs would be willing to pay players after two seasons in the second division.

Antonio must calibrate and communicate the margin of error to the Federation: there are cases of false positives, i.e. players who were thought to have performed well on the national team after two years in European leagues and who did not improve their performance, as well as the opposite, i.e. false negatives: players ruled out for participation in youth and national teams, who reached the European market without the support of the Federation, after having been evaluated by means of individual assessment tools. In

one of the meetings, Antonio reports that the number of false negatives remains stable, but that the number of false positives is increasing, for which he must reassess the application.

References

- Alamar, B.** (2013). *Sports Analytics: A Guide for Coaches, Managers, and Other Decision Makers* [English edition]. New York: *Columbia University Press*.
- Charan, R.** (2007). *Know-How: The 8 Skills that Separate People who Perform From Those Who Don't*. Cornerstone.
- Davenport, T.** (2014a). *Analytics in Sports: The New Science of Winning*. SAS. Retrieved from: https://www.sas.com/en_au/whitepapers/iia-analytics-in-sports-106993.html
- Davenport, T.** (2014b) *Big Data at Work: Dispelling the Myths, Uncovering the Opportunities*. *Harvard Business Review Press*, [240]. USA ©2014.
- Ford, P.** (2015). What is Code? Bloomberg Businessweek. Bloomberg
- Hohman, A., Lames, M. and Letzelter, M.** (2005). *Introducción a la ciencia del entrenamiento*. Barcelona: Paidotribo.
- [Untitled image of Benjamin Alamar]**. (undated). Retrieved from: https://www.linkedin.com/in/benalamar?trk=public_profile_browsemap_mini-profile_title.
- [Untitled image of Paul DePodesta]**. (2017). Recovered from [https://commons.wikimedia.org/wiki/File:Paul_DePodesta_\(36380255290\).jpg](https://commons.wikimedia.org/wiki/File:Paul_DePodesta_(36380255290).jpg).
- Jayal, A., McRobert, A., and Oatley, G.** (2018). *Sports Analytics (Routledge Studies in Sports Performance Analysis)*. London: Routledge.
- Lebed, F.** (2017). *Complex Sport Analytics (Routledge Research in Sport and Exercise Science)*. New York: Routledge.
- O'Donoghue, P.** (2014). *Data Analysis in Sport (Routledge Studies in Sports Performance Analysis)*. London: Routledge.
- Whitehead, A.** (1967). *The Aims of Education and Other Essays*. New York: The Free Press.