

# WHITEPAPER: Using Data to Drive Sport Development

March 2019



## Executive Summary

The days are over when high performance sport can rely on precocious athletes to consistently rise to the top of highly segmented organizational structures. National sport governing bodies (NGBs) worldwide are shifting their focus to developing what many term the 'grassroots' level of their programs knowing that some of these young athletes will eventually rise to the top of the national rankings. Future performance is built upon youngsters having opportunities to learn, practice, and compete and NGBs are designing programs to help this happen.

Making informed decisions requires data, and lots of it. NGBs should have access to the following kinds of data:

**Membership data describing the athlete pool.** Knowing who the athletes are, where they live, how old they are, and whether they are boys or girls is the most basic kind of data to collect. All other data sets are based on athlete membership figures and member statistics.

**Retention data** to determine how long athletes (of all ages) have been involved in NGB programs. Time is the most important factor in talent development so knowing retention percentages is necessary for informed evaluations about programs and clubs.

**Growth data** to help coaches design effective teaching and training plans. Simple, periodic measurements can be used to predict an athlete's *age at peak height velocity* (APHV), an important training milestone.

**Administrative data** that describes the associations training and certification of officials and coaches.

**Performance data.** Track performance over the entire span of an athlete's participation in the sport. This kind of data is used to determine athlete rankings, evaluate training protocols, and

---

provide motivational information for athletes, parents, fans, media, and others who may be interested in long-term performance information.

A data system will help NGBs make decisions with facts and numbers rather than guesses based on anecdotes or tradition.



## About this report

This whitepaper is published by Sportkid Metrics LLC, which is responsible for its content. Questions about its content or about Sportkid Metrics should be directed to:



Sportkid Metrics LLC  
<https://sportkid.asia>

Author: William J Price  
price@sportkid.asia

Contact: Phone: +60 17 697 3951 (Malaysia)  
570 955 3526 (United States)

Issue date: 1 March 2019

© 2019 Sportkid Metrics LLC

# Introduction

Traditionally, sport development has been addressed as though coaching expertise, sports minded administrators, and capable and willing athletes were all that were needed to achieve national and international success. Success achieved through these factors alone though has been limited to countries with large populations, accessible facilities, and what has vaguely been referred to as sporting culture. Smaller or developing countries that relied only on these factors do not fare well internationally.

In the early days of international sport the size of the athlete pool, the number of coaches and available facilities did indeed matter, but not anymore or, at least, not as much. Science now determines progress in sport and both performance data and demographic data are essential to modern sport development. While training science is important, especially at elite levels, knowledge of the demographics and infrastructure affecting sport is vital to raising performance and building sport overall.

Data allows coaches and administrators to abstract the development process. Instead of focusing on single performances or specific athletes data provides the 'big picture' of what is happening within the sports' domain. This includes athlete growth and retention, coaches and officials training, membership demographics, and distributed governance by state or local branches of national bodies.

There are technical details involved in both collecting and sharing data, however they are two different activities and need to be considered separately. Collecting data is more difficult primarily because a strategy has to be developed as to how, who, where, and when data is collected. Compared to collecting data, sharing it is easy.

## The need for data

Modern efforts to develop sport must rely on data and statistics rather than guesses. Good decisions are based on facts; experience and past practice are important but they are not substitutes for facts. Decisions based on what someone thinks is right or using past practice as a path to future success often leads to poor outcomes.

Experience is based in tradition and anecdotal evidence, which does not stand up to real world challenges. Sometimes NGBs, clubs, and coaches do things a certain way because that's the way they have always done them; inheriting structures and procedures from previous eras, the efficacy of which have never been reviewed.

Good data is vital to sport development. Most NGBs are good at collecting data from their high performance programs possibly because there are fewer athletes to deal with, thus making data collection easier. Grassroots initiatives are different and they present challenges of scale not found at the high performance level. There are many more athletes involved and the systems required to properly track and analyze these programs often don't exist, are poorly utilized, or suffer from design flaws.

This whitepaper will discuss several key metrics in sport development and how NGBs can use them to make informed decisions. It will also examine the differences between the kinds of data needed at the various developmental stages.

## Definitions

For the *purpose of this whitepaper* the following terms and abbreviations will be used:

- **National governing body (NGB)** - A national governing body is a country's single administrative body for a particular sport on the national level. Each sport (or group of sports, such as aquatics) will have its own governing body. The NGB, as the term suggests, *governs* the sport within the country through administrative and regulatory procedures.
- **Athlete pool** - This refers to *all* athletes who participate in a particular sport regardless of their level of participation or performance. This includes grassroots level athletes all the way up to national level athletes. The athlete pool has to be understood in a comprehensive manner to track growth of the sport over time.
- **Athlete retention** - The retention metric is the NGB's most important statistical indicator of long-term talent development. Retention is reported by percentage and is a measure of how many athletes remain involved in NGB activities from one period to the next.
- **Age at peak height velocity (APHV)** - Peak height velocity is a significant growth period usually lasting between 12 to 18 months where an athlete is growing at the fastest rate they will ever reach outside their first year after birth. This period is commonly known as the adolescent growth spurt. It is believed that the attributes of endurance and strength are more sensitive to training effects during this period; thus, knowing when this period is occurring can help a coach design effective training strategies to take advantage of these changes.

## Demographic and performance data

Two basic types of data dealing with demographics and performance are needed by NGBs. Due to privacy concerns demographic data is used internally by the NGB and shared strategically with necessary administrators and identified users of the data system. Performance data is publically available.

**Demographic** data primarily describes the athlete pool:

- Sex and age
- Location
- Retention information
- Nature of participation

Demographic data provides information on the growth of the athlete pool and creation of future talent within the sport by understanding retention percentages and the nature of an athlete's participation. This is called *population* data because it represents a general description of all the athletes in the sport.

Tracking athlete growth can help coaches deliver sound training programs. With periodic measurements performed by local club coaches an athlete's growth history can be used to calculate their predicted **age at peak height velocity** (APHV), an important training metric, and **predicted adult height**, an important attribute in many sports. Data specific to one athlete is useful to the athlete and coaches in setting or evaluating individual training protocols.

Demographic data is also used to quantify expertise and facilities; tracking the number and training levels of coaches and officials, for example, or mapping of facilities in both rural and urban areas. While this may fall outside the meaning of *demographic*, tracking infrastructure and expertise helps an NGB understand the scope of assets available for developmental efforts.

Having a full demographic picture of the sport can make a significant difference in how development is approached. If facilities are adequate then development decisions can focus on how to introduce new programs. If facilities are inadequate or don't exist in particular areas then planning will revolve around infrastructure rather than programming.

**Performance data** is what most people think of when they think of sport. This data describes athletic performance in competition, test events, or training environments. It is valuable not only in tracking athlete improvement but also in designing motivational schemes, promoting the sport in the media, and establishing milestones throughout a sports history.

Performance data should be available online and searchable not only for administrators and coaches, but also for the general public. Providing this kind of information to young athletes, their coaches, and their parents creates a powerful motivational tool for the sport. It also creates transparency between sport administrators and the public who may appreciate an accounting if public money is spent on sports.

NGBs should never discount the motivational value of having good information available to all involved. Once people begin to understand the bigger picture their involvement starts to take on a different tone and they are more likely to buy into national initiatives. They can see where they stand in comparison to other athletes, coaches, and clubs.

## Considerations and methods

For many years sport bodies have applied the typical sport sciences such as physiology, psychology, nutrition, and pedagogy to the development of athletes and training of coaches in high performance programs. These programs utilize science with laser-like focus to help high level athletes perform at the edge of their ability.

But at the lower levels of sport development, what many term the 'grassroots' level, the sciences are different and based more in child development, demographics, sociology, and in the broader areas of economics and education. Science used in high performance is of little use at this stage of the sport. NGBs are shifting their focus to social sciences to help them grow developmental programs, create talent, and produce the high performance athletes of tomorrow.

There are two primary differences that affect the data requirements of high performance and grassroots programs:

- **Scale** - High performance programs are small. There can only be a limited number of athletes considered the 'best' in the nation. This smaller number makes it easier to understand what data is needed, easier to collect, and easier to analyze.

Successful grassroots programs have many thousands of participants. Even though the kind of data collected is different at the grassroots level the sheer size of the programs makes such collection difficult and time consuming. NGBs often don't have adequate systems in place to collect and analyze grassroots data.

- **Purpose of the programs** - The purposes of high performance and grassroots programs are fundamentally different: High performance programs *maximize* talent; developmental programs *create* talent. High performance programs focus on getting a select group of athletes to perform at the highest level possible. Developmental programs train thousands of youngsters in sport skills and gradually develops athletes to higher and higher levels in the hope that some of those athletes will eventually join high performance squads.

The different purposes require different kinds of data. As mentioned previously, developmental programs depend more on demographic information. Getting a clear picture of the demographic landscape provides an NGB with the information it needs to make informed decisions. In developmental programs understanding the size and nature of the athlete population is more important than performance.

**Grassroots development is the caterpillar to the high performance butterfly.** One is an easily identified, big deal activity; the other not so much but both are essential to successful development of national sport. Neither can exist without the other and both are essential NGB functions.

### Inclusive membership

Adopting a concept of *membership* in the NGB for all athletes, coaches, officials, and administrators makes it much easier to determine and administer the demographics of the NGB. There are several ways to implement such a concept and NGBs should create and

maintain an inclusive membership process by establishing criteria and compliance regulations. Additionally, once defined, membership in the NGB must be required. Criteria for membership should be set for all roles (coach, official, etc.) but as a minimum guideline *all athletes must be members*.

NGBs not currently operating as membership organizations should consider implementing such a concept. There is no defined time period for any membership scheme to be put in place but the overall goal should be a full membership organization at some reasonable future date.

## System characteristics and NGB demographics

Describing the athlete pool may be the most important aspect of any data system but it is not the only kind of data needed. NGB administrators must have a clear picture of the state of the sport. This includes the following:

1. **Athletes** - Understanding the athlete pool is the first step to creating intelligent developmental strategies.
  - Who are the athletes and where do they live?
  - Are they members of clubs?
  - How old are they and are they boys or girls?

This kind of information forms the basis of the data system. Every athlete that takes part in any NGB program, competition, or training session should be registered as an athlete. Sport demographics and performance information is based on the athlete pool.

2. **Officials** - Adequate numbers of officials need to be trained. Who and where the officials are is important to local and regional development.
3. **Coaches** - Similar to officials, NGBs need to have a registry of coaches, their level of training or certification, where they live, etc.
4. **Clubs** - NGBs need to know where their clubs are, what facilities they use, who their coaches are, how many athletes they are training, etc. Clubs are the lifeblood of a sport and NGBs need to be able to support them but first they need to know where they are and the details of what they offer.
5. **Facilities** - NGBs need a clear picture of a sports' infrastructure to administer programs, training, and competitions effectively.
6. **State or local associations** - Subdivisions of the NGB are key components in the administration of the sport. These bodies play vital roles in development programs and their effective governance determines the overall success of any NGB initiatives.

## Distributed data entry and database design

Collecting data on a national level is a large task. To be effective a data system should be able to distribute the collection of data over among multiple users. NGBs can organize collection efforts in ways that suit their organizational structure. The most obvious distribution scheme is geographic with local associations or divisions of the national body collecting data for association athletes and members within their geographic boundaries. Other methods divide data collection based on roles such as club administrators, coaches, and officials. But no matter how it is done the distribution of the collection and entry of data leads to faster implementation of the system and more timely generation of useable information.

The primary benefit of an **online information system** is the facilitation of the data collection process. A common perception is that large data systems are complex and cumbersome. But a sport development system is conceptually simple. The benefit to such a system includes:

1. Easy access to membership data for athletes, coaches, officials, administrators, and clubs. Having this information in a single location, available to anyone who needs it, is worth the cost of the system itself.
2. Data collection has advanced to the point where database interfaces are now available for general use online. **This makes data entry easier** because the task can be distributed amongst various users.
3. The usefulness of the system depends on the **initial database design**. Users of the data system will only notice design when they are creating reports and retrieving data. A good design of the database internals will offer simple user interfaces and rapid production of reports and other data. A well designed database is easy to maintain and as needs change it is easy to modify.
4. Reports should be designed with the end user in mind. Presentation of data has a direct effect on how well it is understood. Graphs, charts, and tables should be designed so that data is not only presented correctly but that the presentation helps users understand what the data means.
5. The system design has two levels of user to consider: The first are those responsible for its overall administration, which includes assigning other administrators and local or regional registrars who are primarily responsible for data entry within their state or region. The second level are casual users who do not need specific training on data entry because they use the data rather than create it.

## Growth tracking

Growth tracking can be one of the most useful tools at a coach's disposal. Regular measurement of four key areas (height, weight, arm span, and sitting height) can provide important information about an athlete's growth trajectory.

Equations used to calculate the growth estimates are more accurate if measurements are done periodically; thus, athletes should be measured three to four time per year.

**Determining age at peak height velocity (APHV)** - Peak height velocity represents the fastest period of growth for a youngster and is commonly known as the adolescent growth spurt. Determining when this growth period occurs can aid the coach in making important training decisions. Through periodic measurement the athlete's APHV can be estimated.

These calculations use either the Mirwald or Moore *maturity offset* equations, which, when subtracted from the athlete's current age, predict APHV.

**Predicted height** - An athlete's full adult height can be predicted if the height of both biological parents is known. Parent height only has to be measured once but this measurement should be as accurate as possible to get good estimates.

**Percentage of predicted height** - The athlete's predicted adult height doesn't change after the initial parent heights are entered but the percentage of the athlete's predicted height, when evaluated in conjunction with APHV, offers a good indicator of growth. Percentage of predicted height is calculated following each measurement period.

**Arm span ratio** - The ratio between arm span and height is an important metric for some sports (basketball, volleyball, boxing, and others). In adults the ratio is usually close to 1.00 meaning that the person's height and arm span are pretty much the same. Variation is common and athletes with higher ratios (1.03+) may be especially suited for specific activities or positions on their teams, or they may enjoy certain advantages in their chosen sports. Young athletes show large variation in arm span ratios so this measurement is not meaningful until adult height has been reached.

**Body mass index (BMI)** - Properly used, BMI is a population statistic not an individual one, however BMI can be used as another growth indicator for athletes.

## Athlete retention

Time is the most important factor in talent development, thus knowing how long athletes have been involved in sport training programs is an important metric. The athlete membership database can help coaches, clubs, and national associations track retention.

Keeping athletes involved in sport *long enough to make a difference* is a key retention objective. Tracking retention numbers without a membership database is almost impossible because it relies only on anecdotal reporting from coaches and clubs. With a proper data system the retention percentage can be accurately reported.

With accurate retention data national associations can modify programs, competition formats, and make other changes to keep youngsters involved in the sport. They can also act to strengthen the stability of clubs and coaches in regions around the country.

**Retention is the single most important metric when it comes to overall sport development.** It provides information about the current athlete pool and how many of those current athletes continue their participation from year to year. It's important to track this metric and to make sure that coaches and administrators know what it means and that they understand its impact on the developmental process.

## Using the retention metric

Reported as a percentage the retention metric represents what portion of an athlete cohort continue its sport participation from one training or tracking period to another. A typical tracking period is one year because it's set to coincide with NGB membership re-registration.

For example, if 100 athletes register for training in 2019 and 70 of them re-register for the sport in 2020 then the 1-year retention rate is 70%. If 65 of the original cohort re-register for the sport in 2021 then the 2-year retention rate is 65%, and so forth. Athletes are tracked within the cohort they originally started in.

If tracking retention is a new thing for an NGB then the first few years of tracking will establish the current norm. Determining what a good retention rate is will probably take several years of data. However, tracking retention of clubs or local associations will give a quicker indication of good versus not so good retention rates.

Once the NGB becomes familiar with what their normal retention rate is they can begin creating ways of raising it and investigating why athletes leave the sport. Since successful talent creation depends on keeping the largest number of athletes involved for the longest time the retention metric is an important administrative measurement. Athletes will drop out of sport programs for a number of reasons. The goal of an NGB is to discover and mitigate these reasons as much as possible.

## Conclusion

Sport governing bodies are looking for ways to improve national and international performance of their athletes. After a period of using sport science to fine tune high performance athletes administrators have now set their sites on improving developmental programs that feed talented youngsters into national programs. To do this attention is shifting to the kind of information that NGBs need to build up developmental programs.

Comprehensive collection, tracking, and analysis of demographic and performance data is necessary so that NGBs can properly administer and grow their sports. Data systems for this task can be professionally created for relatively low cost. A well designed system requires minimal training and any training would focus mainly on interpreting the reports generated by the data system.

Using data in this way is new. With it administrators can expand their understanding of what development means. When that happens possibilities that have never been seen before suddenly appear.

Making decisions based on guesses and past practices is no longer an acceptable or effective way for NGBs to operate, especially if public funds form part of an NGB's operating budget. Data provides facts that inform decisions.

