The Role of Databases in Sport Science: Current Practice and Future Potential

Jenny Vincent, Pro Stergiou, Larry Katz

Faculty of Kinesiology, The University of Calgary, Canada

Abstract

Databases with the tools for capture, storage, management, retrieval, integration, analysis, interpretation, reporting, and dissemination have the potential to be the single most powerful tools in sport science. Knowing how to collect, store, access, retrieve, and integrate information is critical to effective performance analysis and decision-making. Databases should form the underlying foundation of most other tools used in sport science as they provide the structure and access to the information that is the catalyst for most other applications. As the development of resources become more integrated, the value of databases increases and their role in system design becomes essential. This includes the ability to retrieve previous data for comparison with new performance and the use of data to highlight issues for deliberation. Databases are also useful as multimedia repositories of sports information. Sport involves human movement. This can be quantified with numerical data, graphics, and audio/video recordings. Multimedia resources in this context are exceptionally valuable, especially if the information can be accessed via creative interfaces that provide timely and efficient information that is tailored to each individual's unique requirements. This paper discusses the nature of databases, the role of databases in sport science including design considerations, issues of integration, examples, and future potential.

Understanding Databases

A database, by definition, is an integrated collection of logically related records or files consolidated into a common pool that provides data for one or multiple uses. At the simplest level, creation of a spreadsheet in Microsoft Excel constitutes a database. At a more advanced level, databases can be complex and customized systems that capture, store, retrieve, interpret, report, and disseminate information.

Databases form the underlying foundation of most other tools used in sport science as they provide structure and accessibility to data that is the catalyst for many other applications. This structure and access allows the user to retrieve previous data for comparison with new performance information or allows him or her to use data that highlights issues for deliberation. Sport involves human movement that can be quantified with numerical data, graphics, and audio/video recordings. Multimedia resources in this context are exceptionally valuable, especially, if the information is accessed through creative interfaces that provide timely and efficient information tailored to an individual's unique requirements. As the development of these information resources becomes more integrated, the value of databases increase.

A database is essentially a collection of interrelated data of different types. Databases for the

storage and retrieval of information have been very popular in the field of Epidemiology (the study of the factors that affect health and wellness in a population). An example of such a database is the Injury Surveillance System (ISS) that is maintained by the National Collegiate Athletic Association in the United States (NCAA) (Dick, Agel, & Marshall, 2007) and used for tracking of injury data in college sports. This database collects injury data from NCAA institutions in a wide variety of sports and has been in operation since 1982. The type of information collected and stored includes data such as injuries sustained by athletes, the school the athlete attends, and the sport in which they participate. By capturing this data year after year, decision makers use this information to look at information aggregated by institution, sport, or type of injury. It has been used in tracking injury rates in new sports, on different playing surfaces, and to follow the effects of rule changes in various sports.

Database Design

Databases can be used to store all different data types including continuous or categorical data. As related to sport performance, this data can include performance data from competitions or any other type of data that a coach collects during the pre-season, competitions, or post season. Furthermore, this data can include specific biomechanical, physiological, or medical data related to each athlete. Proper planning, creation, and maintenance of a database are required to create and utilize an "active" tool that can benefit all participants. Complex, multiuser databases usually require a database expert to create and maintain.

Interpreting information from a database is only as good as the initial information that goes into it. This information must be accurate with checks in place to ensure the data is valid. This quality assurance is critical because information frequently comes from a variety of sources. Data can be input manually or automatically. Manual entry requires more time, but ultimately brings a level of confidence that the data is more accurate if the proper checks and balances are in place. Scripts (programs) can be written by a computer programmer that automatically read data, for example, from a website that contains performance data. However, using these programs still requires manual checking to ensure the accuracy of the information that has been automatically input.

A structured and thoughtful database design is essential to the success of a database. One of the key principals of good database design is avoiding unnecessary redundancy. In the Injury Surveillance System (ISS), this principal might be highlighted by the following example. All the data about an institution (for instance the name, abbreviation, and location) is not repeated for each of the 1.4 million injuries collected in the database. The institution information is stored once and then for each injury a 'referential integrity constraint' is created to link to the institution information. By avoiding the duplication of the institution information the institution information is kept consistent allowing successful retrieval and aggregation of injuries across institutions. Therefore any updates to the institution information need only be made in one location.

Another principal of database design is ensuring the uniqueness of information. For example in the ISS it is critical that every injury is recorded only once. An injury may be uniquely identified by the date and time it took place, the athlete's institution, the sport, and the body part injured. These pieces of information make up what is called a key for the injury data (note: since athlete identification information is not recorded in the ISS it can not be used in the key). The key is used in the database to enforce uniqueness. In the injury database example the key is made up of many pieces of information, a database designer will often substitute this for a single computer-generated identification number (e.g. injury id) called a surrogate key.

In any database it is very important to impose validation conditions on the data as it is entered. This is often accomplished using domain integrity, that is, by providing a list of possible choices. Domain integrity ensures consistency of the data maximizing the ability to search and aggregate the information. This type of validation condition is used heavily in the ISS and has attributed to the system's success and longevity. For example, in the collection of data on the principal body part injured a list of body parts is provided. Without this domain integrity a broken toe might be recorded under the principal body part foot, or toe, or even metatarsal making it impossible to access accurate injury rate information on this type of injury. Another type of validation condition is the designation of certain pieces of data as required (i.e. not null). Null values are a representation of missing or unknown data. A careful decision must be made for each data field that will be collected, balancing what information is needed for retrieval and aggregation with what is readily available at data entry time. A final type of validation condition is ensuring the type of data (numeric, character, date, image) and the maximum size, are appropriate for the data that will be collected in each data field. This is imperative to ensure correct sorting, grouping, and retrieval. The thoughtful consideration of all these database design elements will ensure the integrity, usability, and longevity of the database.

The principals of good database design noted above are derived from the process of normalization created by E. F. Codd, who was the first to describe the concept of relational databases (Codd, 1970). Although other database models exist (including hierarchical, network, and object) the relational database model is widely used, is very mature and the database management systems which implement it are fast and reliable. Relational databases are based on the concept of tables with columns and rows. Tables can be related to other tables through keys. Information is retrieved from tables using the Structured Query Language (SQL). In this paper all references to databases are specific to the relational database model.

Database Management Systems

A database management system (DBMS) is a set of computer programs that control the creation, maintenance, and use of a database (Wikipedia, 2009a). That is, it allows the designer to create tables with columns of a specific data type and size, and relationships between the tables; it allows for the insertion, update, and deletion of rows in the tables; and it allows for the retrieval of information from the tables using Structured Query Language (SQL). SQL is a database computer language designed for managing data. Two of the most commonly used desktop DBMS' are Microsoft Access and Apple's Filemaker Pro (cross platform). In multi-user DBMS environments, Microsoft SQL Server is a popular choice. The industrial strength Oracle DBMS is often the preference for large organizations and data warehousing applications. The freely available MySQL is another popular alternative amongst database developers.

Content Management Systems

A content management system (CMS) is a computer application used to manage workflow needed to collaboratively create, edit, review, index, search, publish and archive various kinds of digital media and electronic text (Wikipedia, 2009b). Most systems use a database to store content, metadata, and artifacts used by the system. The database will often manage security of the content allowing users to view, insert, update, or delete content as specified. The most common type of content management system is a Web CMS, which manages web site content. It will provide authoring tools to allow users with no knowledge of web site creation to maintain parts of the site. One example of a Web CMS in the sports world is a product used by

hundreds of university athletic departments across North America called SideArm (Internet Consulting Services, 2009). It is used to provide information about a university's varsity sports teams allowing the team coach to easily update information as the season progresses.

Sport Databases

Databases are used in sport administration, in all of the sub-disciplines of sport science and in sport epidemiology). Examples from some of these areas are provided below together with a discussion of the key elements of good database design.

Sport Administration

An interesting possibility for databases within the sports administration stream is in the scoring of athletes competing in judged sports. A recent paper describes a gymnastics scoring system, which separates the video performance of an athlete into a set of component parts or skills and compares each skill to a database of skills and the scoring of the skill as defined in the rule book (Jeongeun Shin & Ozawa, 2008). In the rulebook for the gymnastics horizontal bar discipline, 141 skills are described and key poses for each skill displayed along with skill classification and difficulty level. In the skills database a silhouette image of a 3-D human body is augmented with the posture parameters from the key poses. Using image processing the athlete performance is then compared with the 3-D key positions from the database to provide scores for each skill and then the whole routine.

Epidemiology of Sport Injury

As noted above, epidemiology is a data intensive field of sport science. At the 2008 Summer Olympic Games in Beijing an injury surveillance system was used not only to track epidemiological information, but also to direct injury prevention programs (Junge et al., 2008). The system was shown to effectively track the large multi-sport events at the Summer Olympic Games. The researchers provided all definitions, forms, and analysis details so that the injury surveillance system could be implemented at other sports events, thus growing the overall base of standardized data. This type of work allows for comparison across various sporting competitions as well as supporting longitudinal studies nationally and internationally.

Databases in Performance Analysis

One of the most widely available types of sport performance data is competition result information. The National Performance Database is one example of a recently created database that manages performance competition data. This database was created by the Sport Technology Research Laboratory (University of Calgary) for Own The Podium 2010 (OTP) podiumcanada.ucalgary.ca:8080/OTP). OTP 2010 "… is a national sport technical initiative designed to help Canada's winter athletes win the most number of medals at the 2010 Olympic Winter Games in Vancouver, and to place in the top three nations (gold medal count) at the 2010 Paralympic Winter Games" (Own the Podium, 2009).

The National Performance Database includes competition results and performance data for all 152 sport disciplines participating in the 2010 Winter Olympic and Paralympic Games. This database includes competition results from all the World Cups, World Championships, and Olympic Winter Games since 1997 (Vincent & Childs, 2009). Using a web-based interface, administrators and the public can search for results by the various competitions or by an athlete's name. Also, the National Performance Database provides a Canadian medal count and a competition calendar. Decision-makers at OTP have a number of other interfaces that allow them to analyze Canadian athlete performance, compare Canadian results, and even view

the medal counts of the top 10 countries. Moreover, the database offers a comparison of individual athlete performances to their benchmarks, as well as, providing the OTP decision-makers with information on which countries are the strongest in various sports and disciplines. All data is live and new competitions are loaded as they occur on a weekly basis. The data is presented in both a tabular format and as well as graphs.

Figure 1 shows a sample extraction from the Own the Podium National Performance Database for short track speed skating.

Figure 1. A query of the National Performance Database that shows information related to short track speed skating.

There were a total of 38 medals won by Canadians in 2008-2009: 9 gold, 10 silver, and 19 bronze medals. Users can drill down further by clicking on the "+" sign to the right of the word Short Track Speed Skating to see in which disciplines and genders Canadian athletes won the medals.

Another area where performance analysis databases are effective is related to physiological and/or biomechanical data. If the data is properly organized, then online database tools such as SSENTIF Sport (http://www.ssentif.co.uk). have the capabilities to collect, manage, monitor and analyze the data providing valuable tools to improve athlete performance with an interface that can be used by those without database expertise. Tables and graphs can be generated to track athlete progress and development. In the following example, biomechanical data related to how fast an athlete can throw a ball was collected including ball speed and body positions of the athlete as they performed the skill. All data was entered into the database and graphs were generated on how the athletes performed. Figures 2 below shows example graphs from two athletes, one experienced and one novice.

Figure 2. Data graphs generated from SSENTIF software. The image on top is a developmental athlete and the image on the bottom is an elite athlete.

Other programs such as Qlickview (http://www.qlikview.com) provide tools for the novice database users so that they have the opportunity to explore the data, generate visual representations of the information, and hopefully improve decision making using evidence based considerations.

Video Databases

In the analysis of sport performance, video remains a valuable element to success. Video analysis is used by both the coach and the athlete to improve skill development, reduce injuries, provide tactical information, and support notational analysis. The automatic indexing and retrieval of specific video content from full competition film is an ongoing research area. The SportsBR system is able to take in a video from the massive number of sporting competition recordings available, automatically select event-based clips from the video, and index and store them into a database for later browsing and retrieval (Hua-Yong Liu, 2005).

The system automatically selects events by looking at audio, visual and caption text information. It then indexes the clips using speech recognition technology and the extracted caption text. This integrated use of different methods to select clips from a larger video, termed multimodal analysis, is a hot spot in video analysis research. For this research they used soccer matches, selecting the events: penalty kick, free kick next to the goal box, and corner kick. Testing against a set of World Cup soccer matches, they were able to achieve an accuracy rate of 91.3% on clip selection and 97.5% on the subsequent recall of the clip. The keywords used to query the database were derived using the broadcast commentary, via speech recognition software, and from extracts of the captions appearing during the clip. This information was normalized against a sport video vocabulary database created by the researchers from observations of matches and provides some domain integrity to improve retrieval of the clips.

The research into multimodal video retrieval is not only producing databases of video clips with keywords; the research is helping to create storehouses of sports performance data. For example, the GNAVI system is focusing on the capture of the detailed information from the caption text of golf match video thus creating a collection of hole-by-hole player information (Cheolkon Jung & Joongkyu Kim, 2008). These types of systems, along with existing competition results, notational, and physiological datastores provide a mass of data available for analysis. The next step is developing the ability to analyze and extract useful information from this data.

Traditional data analysis techniques, including aggregation queries and statistical analysis, are used and researched frequently in sports. These techniques are hypothesis-driven, that is, they answer a specific question. Data mining is the process of extracting patterns from data; it is a knowledge discovery approach (Wikipedia, 2009c). Shoeman et al (2006) attempt to answer the research question of "whether proven business data mining techniques can be applied to sports games in order to discover hidden knowledge". They conducted a field study in which they data mine South African National Cricket team match information using the industry standard 'Cross-Industry Standard Process for Data Mining' to implement link analysis on the database. The researchers then examine if this type of analysis is suitable for the sports world by presenting the findings to a set of stakeholders of the South African National Cricket team. They found that business data mining techniques and applications do have the potential to discover new information, however, the transformation of match data into a usable format can be problematic. They go on to suggest that the data mining of multimedia databases, containing clips and data automatically retrieved from competition video, might be successful in getting around some of the problems.

More recently another group of researchers also mined the highly available and rapidly growing stores of cricket match data to help coaches determine game tactics and improve coaching strategies based on correlations of play patterns (UmaMaheswari & Rajaram, 2009). These researchers used an object-relational database model to store ball-shots and the varying outcomes. From the database they were able to analyze both the performance of a single player and find patterns for an entire team.

Video Databases in Coaching

Coaches and sport associations habitually collect data in various formats in a quest to improve athletes' performances; and, this data not only concentrates on their own athletes, but also on competitors as well. Managing all the data appears to be a very laborious task that can consume substantial amounts of time. Finding information to facilitate training and decision-making is especially difficult. Learning how to store meaningful information effectively is critical to efficient retrieval and analysis of data. This section discusses databases that

specialize in the storage, management, and retrieval of video data in coaching.

The Digital Scouting system offers wrestling coaches and athletes the opportunity to view and share video over the web (Digital Scouting, 2009). The application allows the user to easily organize and search for video, audio, or other digital resources using unlimited categories, events, or divisions. A permission tool then allows the user to provide a wide variety of access levels including view only, upload, or edit. Video shared through the Digital Scouting site include basic and advanced instructional material, examples of a wide variety of techniques, as well as competition video from the International, Domestic, and University scene. This application provides an example of how databases are used to manage multimedia content and security access to groups of users and various content. A sample of the Digital Scouting web interface is provided in Figure 3 and Figure 4 below.
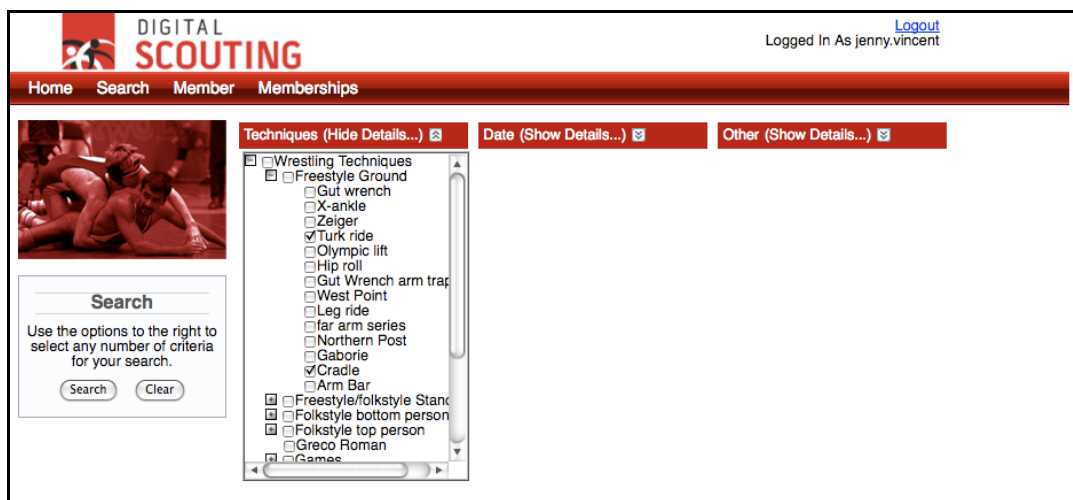
Figure 3. Wrestling video database main screen (Digital Scouting)
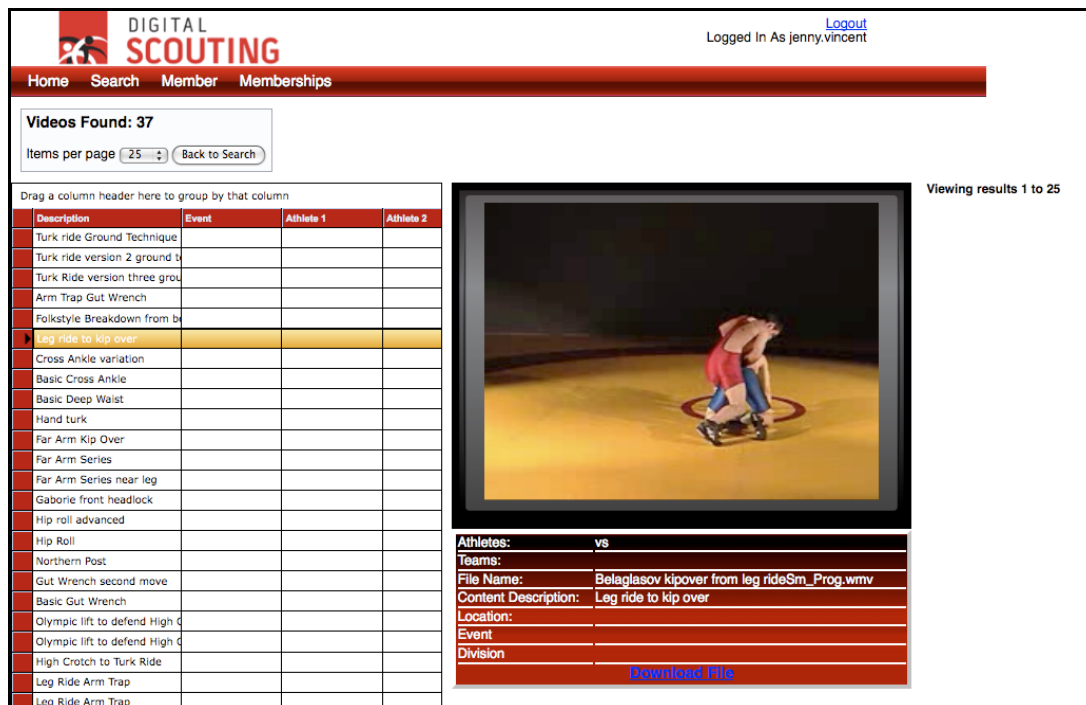
Figure 4. Wrestling video database – Video search display (Digital Scouting)

A similar web-based video sharing application called the Video Report Card has been developed by the University of Calgary's Sport Technology Research Laboratory. This system facilitates the sharing of both video and subsequent comments on the video within the database structure. This is particularly useful for athletes who may have a coach who travels frequently or who lives in another city. An example of a commented video is provided below in Figure 5.

Figure 5. Video Report Card - coach view (Sport Technology Research Laboratory)

An innovative virtual reality system has been built which allows coaches and athletes to see a virtual simulation of a complex series of gymnastic movements (Simin Li & Jinhai Sun, 2009). The coach can design or change a series of movements, looking to maximize the difficulty levels and then watch the simulation to see what the series might look like. A database is used to store the vast amount of data needed to produce the simulations. The database has 3 components: a movement storehouse, a three-dimensional model of the human body, and a movement arrangement storehouse. When realized within the three-dimensional virtual reality environment coaches and athletes can watch the series from all angles. They can then use this knowledge to return to the gym to train more effectively and safely. This example highlights the power of a database to store extremely large amounts of data and then access that data in a

very efficient manner. Today's average databases can easily search and aggregate terabytes of data.

The Web is an ever-increasing dynamic tool with regards to teaching. Sports skills in this setting are no exception. Another application coming out of the University of Calgary's Sport Technology Research Lab provides such an e-learning environment for students who are learning a new skill. The Virtual Coach has the ability to compare an athlete's performance with that of a professional model, thus permitting the athlete to see the subtleties of the skill sets while simultaneously seeing both the model and themselves. Visual feedback, written evaluations and professional demonstrations of the skill sets are used to comprehensively enhance the learning process. The tool adds a common reference point between the athletes and the coaches whereby details of the sports skill can be discussed and confusion eliminated. Having sophisticated resources and training techniques available on the web on a continuous basis can have great benefits to the student and to the teacher. Sample screens of the Virtual Coach interface are shown in Figure 6 and 7 below.

Figure 6. Virtual Coach – Student views self performance and compares to professional performance (Sport Technology Research Laboratory)

Figure 7.   Virtual Coach with demonstration video of skill and online database for self recording of skill performance.

Online video databases provide individuals with the abilities to store, manage and share video content online with other people from around the world. The manager of the database can determine how to organize the videos, the content that will be available for viewing and can manage subscribers to view the content.

Dartfish, best know for its sport video analysis software, has developed Dart Fish TV an online portal which provides video database capabilities (organizes its databases in items called Channels (http://www.dartfish.tv). The image below (Figure 8) shows the front page of a channel called in this example, the Pro Stergiou Demo Channel. The content of what was created and the analysis performed on the user's desktop is up loaded in a simple manner from the Dartfish interface. An image of the process of uploading videos to the database is shown in Figure 9 below. As long as the user has a connection to the Internet, the process of uploading requires just a click of a mouse button. Additionally, content other than videos can be attached to the upload. This might include some performance data that was collected and compiled in a

spreadsheet that the user would like to share with interested individuals. Any type of file can be attached to a video, similar to an attachment in an email.

Figure 8. Image of front page of Dartfish TV website that is used to store, manage and share video content.
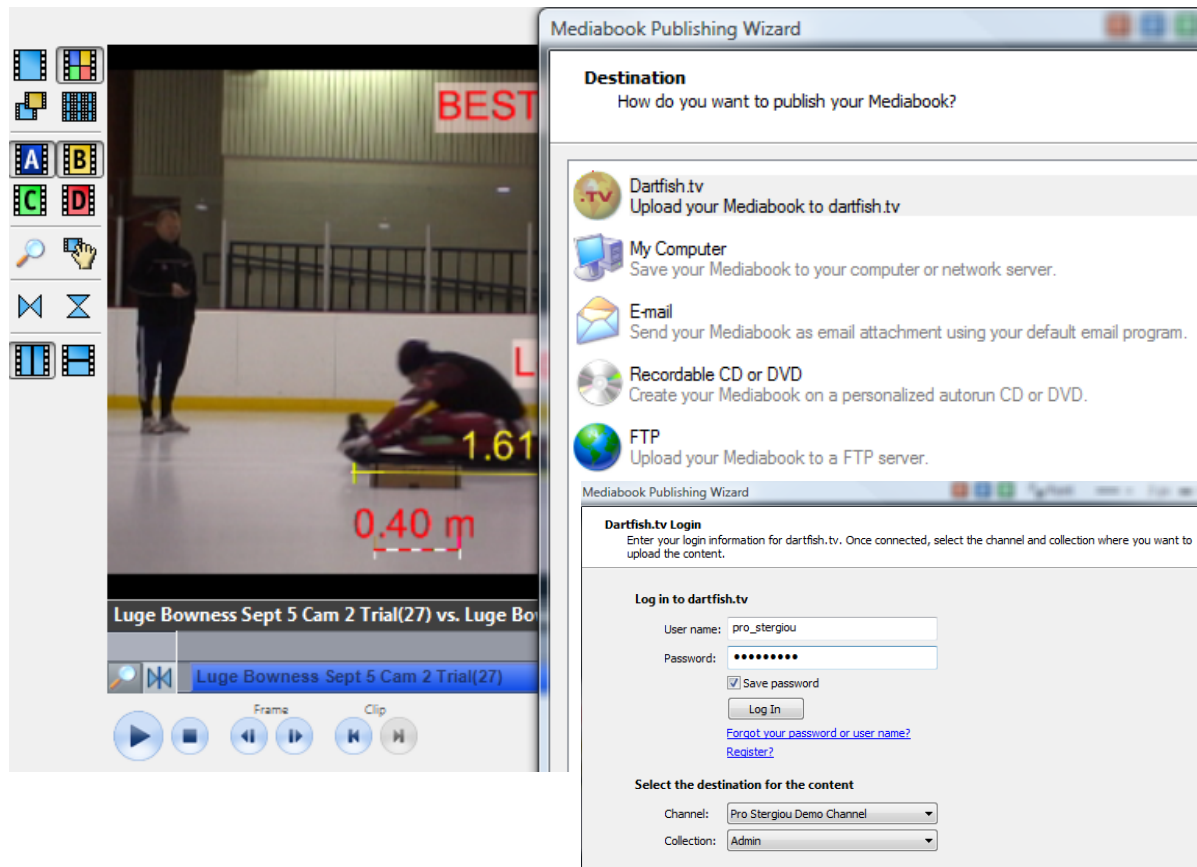
Figure 9. Process to upload video content to Dartfish TV online video database from Dartfish software interface.

Once the videos and the information are uploaded to the online video database, users can access the information from anywhere with the use of a web browser. Access to specific videos and content can be limited to levels of access and the administrator of the channel will create and dictate what content and by whom it can be seen. If analyses were performed on the video (e.g., a breakdown of a movement into key positions) then the person viewing the content would have access to the broken down movement, comments and any overlaid drawings that were placed on the video for analysis. Figure 10 below illustrates a video posted on Dartfish TV showing analysis of the start in the sport of Luge.
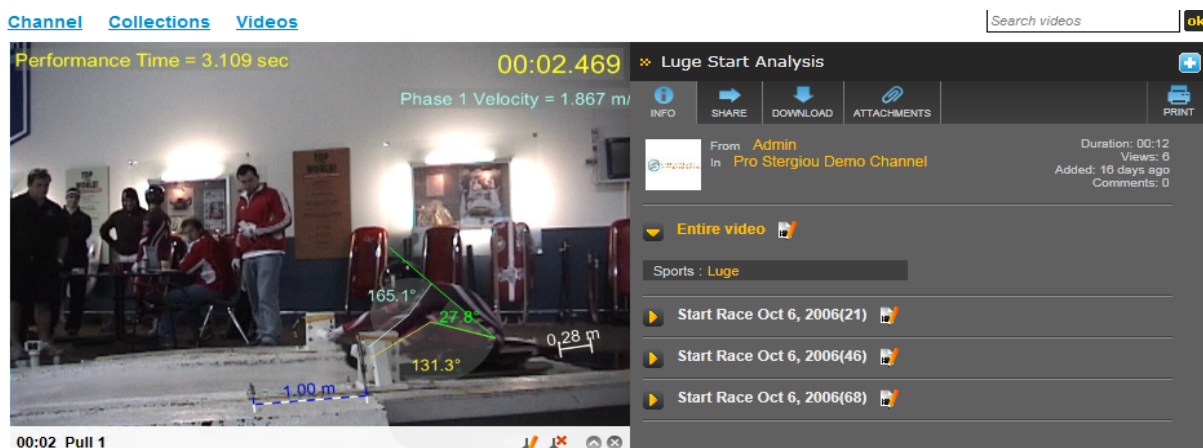

Figure 10. Video posted on Dartfish TV with analysis of the Luge start. The video is shown on the left side with the any text on top of the video. The capabilities to download this video and look at any attachments are shown on top, to the right of the video.

Videos that are related to the analysis of competition/game play can also be uploaded to Dartfish TV. The video, the events and the categorization of the events are all searchable and viewable. Figure 11 below shows a game from a team sport that has been uploaded to Dartfish TV.
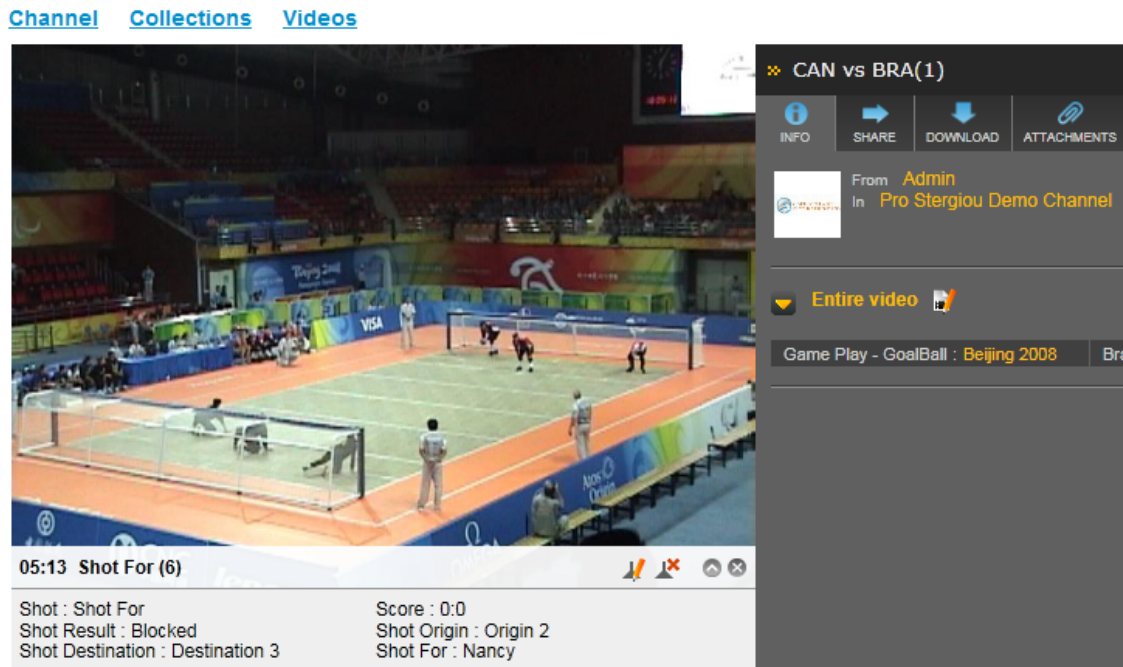

Figure 11. A screen shot of a video of game play during Beijing Paralympics 2008 of the Women's National Canadian Team. The video is playing a single shot from the game, with details located below the image categorizing the shot.

Other publicly available video databases exist for any person to post video content. "You Tube" is the best known one of these sites, however, the capabilities of these types of sites generally do not allow the user to create their own spaces, manage content nor control access. Other types of these online sites exist to store video content (Robinson, 2009; Siegchrist, 2009). Searches on the Internet will provide updated information on other available online video and content management databases using search terms like "uploading video to video sites" or "content management databases for storing sports data".

Future Considerations

The use of multimedia databases has the potential to revolutionize the way coaches, athletes, administrators and the public approach sport. It can impact the way sports are judged, the way athletes are selected to participate on teams, the way coaches approach training plans, and the public perception of sport both at the amateur and professional level.

Many skilled, elite level coaches and sport administrators make decisions based on intuition, 'gut feeling' and the influence of past experience. Effective use of performance databases can encourage sport professionals to employ an evidence based decision-making approach. This does not mean that they need to abandon their 'intuition' rather they need to integrate valuable data modeling into the decision making process. If the data does not support intuition, the data could be wrong, but at least there is a challenge to verify the decision. With the advent of online databases in sport, evidence based decision making is now practical.

For example, it is possible to collect and record experience, performance, injury, and skill data

on each athlete in a given sport. In many countries team selection can be very controversial. In Canada, there is a special organization, the Sport Dispute Resolution Centre of Canada (www.crdsc-sdrcc.ca/eng/home.jsp) whose main purpose is "to help prevent and resolve disputes at the highest levels of the sport system". If sports bodies used evidence-based decision-making more often, it would reduce the number of disputes, as the decisions would be clear and less ambiguous.

However, it is incumbent on those who develop and maintain the databases to ensure that the objects (e.g. video, images, text, skill data) deposited in these repositories be properly organized and tagged so that they can be effectively and efficiently retrieved and analyzed by the appropriate decision makers.

With easy access to information, the issue of data governance is crucial. According to Wikipedia ((Wikipedia, 2009d)

> "Data governance is an emerging discipline with an evolving definition. The discipline embodies a convergence of data quality, data management, business process management, and risk management surrounding the handling of data in an organization. Through data governance, organizations are looking to exercise positive control over the processes and methods used by their data stewards to handle data."

This includes ensuring the quality of the data (completeness, accuracy, timeliness and decision usefulness of the data). Also of critical importance are the privacy policies and procedures that require special attention to storage and sharing of data that are collected on athletes (e.g. medical records). The potential of databases is great; there are many obstacles yet to be addressed; but clearly, it is worth the struggle.

References

Cheolkon Jung, & Joongkyu Kim. (2008). GNAVI: Golf navigation system based on player information*, in proceedings of the 2008 IEEE International Conference on Multimedia and Expo*, Hannover, Germany, IEEE*, June, 23-26, 1413 – 1416.

Codd, E. F. (1970). A relational model of data for large shared data banks, *Communications of the ACM, 13*(6), 377-387.

Dick, R., Agel, J., & Marshall, S. W. (2007). National collegiate athletic association injury surveillance system commentaries: Introduction and methods. *Journal of Athletic Training,* 42(2), 173-182.

Digital Scouting. (2009). *Digital scouting database.* Retrieved August, 2009, from http://www.digitalscouting.com/index.html

Hua-Yong Liu, H. Z. (2005). A sports video browsing and retrieval system based on multimodal analysis: SportsBR*. Proceedings of 2005 International Conference on Machine Learning and Cybernetics,* Guangzhou, China, August 18-2, 8, 5077-5081.

Internet Consulting Services. (2009). *Why SideArm*? Retrieved August, 2009, from http://athletics.internetconsult.com/why-sidearm.aspx

Jeongeun Shin, & Ozawa, S. (2008). A study on motion analysis of an artistic gymnastics by using dynamic image processing - for a development of automatic scoring system of horizontal bar – *Institute of Electronics, Information and Communication Engineers (IEICE)* Technical Report, 108(46), 13-18.

Junge, A., Engebretsen, L., Alonso, J. M., Renström, P., Mountjoy, M., Aubry, M., et al. (2008). Injury surveillance in multi-sport events: The International Olympic committee approach. *British Journal of Sports Medicine, 4*2(6), 413-421.

Own the Podium. (2009). *Own the podium - our initiative*. Retrieved August, 2009, from http://www.ownthepodium2010.com/

Robinson, R. List of Video Sharing Web Sites – Share your Videos, *The Online Video Marketing Guide – News, Tips & Trends,* Retrieved, October, 2009, from http://www.reelseo.com/list-video-sharing-websites/

Schoeman, J. H., Matthee, M. C., & van, d. M. (2006). The viability of Business Data Mining in the Sports Environment: Cricket Match Analysis as Application. *South African Journal for Research in Sport, Physical Education & Recreation, 2*8(1), 109-125.

Siegchrist, G. (2009) Video Sharing Web Sites – Share Your Videos Online, *About.com:Destop Video*, Retrieved November, 2009, from http://desktopvideo.about.com/od/videohostingsites/a/vidsharing.htm

Simin Li, & Jinhai Sun. (2009). Study on simulation of gymnastics movement and the arrangement based on virtual reality technology., ICIC '09. Second International Conference on Information and Computing Science, Manchester England, May 21-22; 3, 270-273.

UmaMaheswari, P., & Rajaram, M. (2009). A novel approach for mining association rules on sports data using principal component analysis: For cricket match perspective*, in proceedings of the IEEE International Advance Computing Conference*, Patiala, India, March 6-7, 1074-1080.

Vincent, J., & Childs, T. (2009). *Database summary.* Unpublished manuscript, Sport Technology Research Laboratory, University of Calgary.

Wikipedia. (2009a). *Database management system*. Retrieved August, 2009, from http://en.wikipedia.org/wiki/Database_management_system

Wikipedia. (2009b). *Content management system*. Retrieved August, 2009, from http://en.wikipedia.org/wiki/Content_management_system

Wikipedia. (2009c). *Data mining*. Retrieved August, 2009, from http://en.wikipedia.org/wiki/Data_mining

Wikipedia. (2009d). *Data governance*. Retrieved November, 2009, from http://en.wikipedia.org/wiki/Data_governance