

Providing Baseball Research at the Touch of a Button

BY TIM WILES

He never played a single day in the big leagues, but Hall of Fame member Dave Smith is one of the most important people in baseball history. As the founder of Retrosheet (www.retrosheet.org), Smith has probably put more baseball information into more hands than any other historian or statistician.

Fans unfamiliar with Retrosheet will be forgiven if they disappear for a few days after their first visit. Users can view the score for every game in baseball history – as well as box scores and play-by-play data for every game dating back to 1957 – and will eventually be able to search even further.

Want to read the box score of the first major league game that your Dad took you to? No, problem – just visit Retrosheet. Perhaps you'd like a detailed list of every triple play in history, or all no-hitters, or every time someone hit for the cycle? Maybe you want a list of "lost home runs," which were eliminated by weather conditions, human error in running the bases or interpreting the rules? Or even data on every player transaction in major league history or a long list of baseball's most unusual plays?

All this is available free of charge, thanks to the devotion and commitment of Smith and more than 100 volunteers, who research the data and double- and triple-check to make sure it's right before it goes on-line.

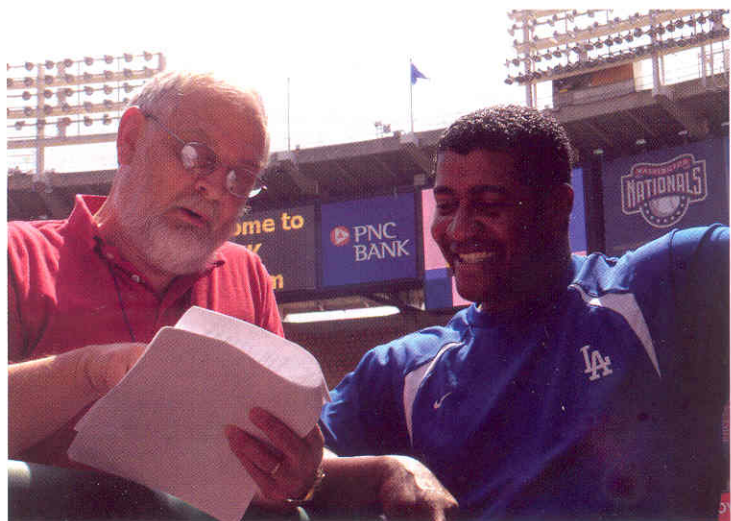
Smith is proud to further his contribution to the world of baseball history by becoming a Hall of Fame member. "That's something that I've been meaning to do for years," he notes. "And I finally got around to it. It's great to support the Hall and all it does to promote and preserve baseball history."

"I was a geeky 10-year-old," said Smith, recalling his own first game, July 18, 1958, at the Coliseum in Los Angeles. His idol, Sandy Koufax, faced only six batters that night, striking out two – but walking four – in the first inning. Smith remembers reading the Dodgers yearbook and finding pages and pages of detailed stats provided by another pioneering baseball statistician, Allan Roth. Smith was fascinated by baseball's ordered, analytical world of statistics. The next year, at age 11, he began collecting play-by-play data.

Twenty-five years later, Smith got an assist from another great statistician, Bill James. James had started a program called "Project Scoresheet," in which volunteers charted all major league games as they occurred, pitch-by-pitch, starting in 1984 and going forward. Many of the volunteers in James' network were intrigued by Smith's idea of also working backward in time, and many of them began to devote time to Retrosheet as well as "Project Scoresheet."

During a visit to the National Baseball Hall of Fame and Museum in 1975, Smith picked up a brochure for the Society for American Baseball Research (SABR). SABR became a rich source of Retrosheet volunteers, who would enter information into a system from scorecards, newspaper accounts, box scores and scorebooks kept by media members and teams. Hundreds of volunteers have assisted over the years, all because they believed in Smith's core principle of providing complex, sortable baseball data free of charge to those who would like to study the game.

In 2005, Smith received the Bob Davids Award from SABR. It is the organization's highest honor, given to an individual whose contributions to SABR and baseball reflect the ingenuity, integrity and



Dave Smith (left) shares a detailed analysis of Mariano Duncan's career with the former Los Angeles Dodgers shortstop, now a Dodgers coach, on May 28, 2006 at RFK Stadium in Washington, D.C.

self-sacrifice of the founder and former president of SABR, L. Robert “Bob” Davids.

At the beginning of the project, Smith encountered resistance from people who considered him an impossible dreamer. Not all the information was available, nor was all the early newspaper data accurate. It would take too long to compile all this information, and no one knew how it would be disseminated. The Web site for Retrosheet went live in 1996, a happy development for someone who had collected data for 12 years to that point.

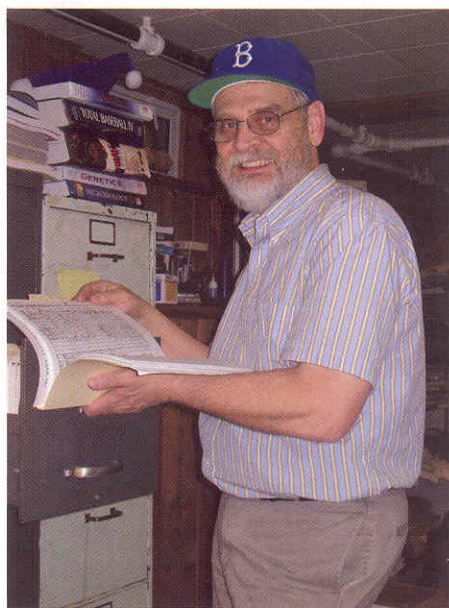
Along the way, Smith developed a relationship with the Museum. “There was one guy there who told me that I was a raving lunatic – this could never be done,” Smith laughed.

But Smith and Hall of Fame researchers began to work together more often, and eventually Smith purchased a gigantic run of baseball’s “day-by-day” data on microfilm from the Library, so that volunteers could cross-check this data against box scores and official score sheets, resolving any discrepancies.

“Dave Smith’s work has made our work here at the Hall of Fame a whole lot easier,” noted researcher Freddy Berowski, who works in the A. Bartlett Giamatti Research Center. “I go to Retrosheet at least a half dozen times a day.” Multiply Berowski’s thoughts by five

research staff members and more than a dozen other library workers and curators, and it becomes clear that Retrosheet has had a major time-saving impact on the Hall of Fame.

Smith, a microbiology professor at the University of Delaware for 32 years, is firmly committed to keeping the site free; he feels baseball history belongs not to any individual or firm, but rather to the nation and its people. “I’ve never made a dime



Dave Smith, founder of Retrosheet, peruses a file of scoresheets from one of the many file cabinets in his basement office.

from this,” he said proudly, though his wife reminds him, “Your hobby is supposed to cost you money – and there are a lot more expensive hobbies than this.”

The site does take donations, and is always looking for new volunteers. It also has a “most wanted list,” where Smith makes a plea for scorebooks and game accounts for which the Retrosheet data collection is currently incomplete.

BHOF ARTIFACT

2006



In 2006, Team USA fell short in their bid to win the inaugural World Baseball Classic (WBC), which was eventually won by Team Japan. Derek Jeter, shortstop for the American team, donated this jersey to the Hall of Fame from the WBC when he batted .450 with nine hits in 20 at-bats.

Of the 189,010 games played since the National Association started in 1871, Retrosheet has play-by-play accounts of approximately 130,000 of them, filling up 17 filing cabinets in Smith’s basement office. “Every major league game box score will be on the site within two years,” Smith confidently predicted. “Especially if voluntarism continues at the same rate – or goes up.”

While a Hall of Famer might be happy with 3,000 hits, Smith’s site often gets more than that on a single day, with a total of 1.8 million since they began counting in 2002. The site is not just used by stat mavens, but also creative types. Actor Billy Crystal was a heavy user while researching the movie “61*,” and director Spike Lee called Smith while studying the career of Jackie Robinson for a possible film project. Some novelists use the site to figure out what happened during a game on a particular day. “We define success not in numbers, but in the happy feeling we get in knowing that we’ve helped people find things out and study the game,” Smith said.

Smith would love to hear from prospective volunteers, and also from anyone with scorebooks or collections of scorecards, especially when kept by a team, an official scorer or a media member. Interested parties may contact him at dwsmith@retrosheet.org.

Tim Wiles is director of research for the A. Bartlett Giamatti Research Center at the National Baseball Hall of Fame and Museum.