

BASEBALL

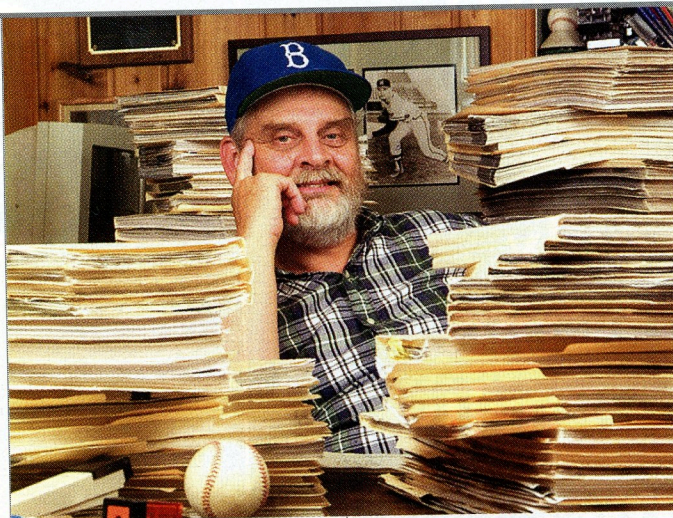
Down to the Last Out

Retrosheet.org is compiling, digitizing and posting the box scores of every game in major league history **By Daniel G. Habib**

LIKE PROUST'S mad-eleine, the box scores at Retrosheet.org encompass little worlds; to sift through them stirs a baseball fan's memories. Anyone who keeps ticket stubs stashed in a dresser drawer or score books stacked in an attic can pass an afternoon this way, meandering through history to find that Cincinnati's Big Red Machine beat the Houston Astros on July 1, 1975, at Riverfront Stadium when Joe Morgan singled home Pete Rose in the bottom of the 15th. (Upon discovering Retrosheet, I didn't wait long before consulting the New York Yankees-Toronto Blue Jays box of Aug. 11, 1994, my 16th birthday and the eve of the players' strike. I recalled sitting in the upper deck at Yankee Stadium but needed prompting to remember that Wade Boggs went 2 for 5 and that Ed Sprague's 13th-inning home run gave Toronto an 8-7 win.)

Retrosheet is immensely ambitious: It seeks to assemble accurate box scores and play-by-play data for every game in major league history, to standardize that information and archive it digitally, and to place it in the public domain for fans, historians and researchers. "We will have proofed box scores with correct totals for every game," says David Smith, a biology professor at the University of Delaware who's been working on the project since 1989. "It's a tough task, but it's absolutely possible to accomplish. Counting all the grains of sand is overwhelming, but it's not infinite. There is an end point."

To date, Retrosheet and its several hundred contributors, all



CRUNCHING HISTORY Smith's diligence has been a boon to researchers and casual fans.

volunteers, have posted box scores for every game since 1969. Data comes from major league teams, all of which have given Smith what score books and records they possess; newspaper box scores; and material from the private collections of beat reporters and ordinary fans. For instance, one Retrosheet member scans eBay for old scorecards, and Smith himself supplied play-by-play of some 80 Dodgers games from 1959 and '60 that he had tracked. All Retrosheet's data is proofed and checked against day-by-day and season totals maintained by the Hall of Fame. Says Smith, 56, "People are disappointed we don't do it faster, but I am paranoid that we get everything right before we put it on the Web."

Smith was a Dodgers fan as a kid in Escondido, Calif. "We went to a game in '58 because Sandy Koufax was my guy," he says. "We went to the Coliseum, and he got bombed, struck out two and walked four, but I

bought a yearbook: red, with the players' signatures and a big baseball on the cover. And in the back there was page after page of detailed statistics."

That data was the brainchild of Allan Roth, a Dodgers statistician whose work presaged much of today's quantitative analysis. "I thought, My God, I want to be him," Smith says. "In some sense, Retrosheet started that night."

It now operates from the basement of Smith's Newark, Del.,

house, which holds 11 file cabinets of data and a DSL server. Not only does the site bubble with irresistible trivia—in the sixth inning on Opening Day 1945, White Sox third baseman Tony Cuccinello retired Cleveland's Lou Boudreau on a hidden-ball trick—but it has also become an indispensable resource for saber-metric analysis. Retrosheet's historic data allows researchers to examine a player's splits (home versus road, lefthanded versus righthanded, average with runners in scoring position), to compile park factors (ratios that express whether a ballpark favored hitters or pitchers), to study a bullpen's impact on a starter's won-lost record and to create dozens of other metrics that are taken for granted in slicing and dicing contemporary data. "It brings the 1960s up to the 2000s in terms of analysis," Smith says.

Smith has not yet uploaded the box from that Dodgers game—July 18, 1958—because Retrosheet is moving backward systematically. (Contiguous sets of data serve researchers better.) His father taught him how to keep a scorecard that day, but he won't use his own record when the time comes. "We have Allan Roth's scorecard from that day," he says with a laugh, "which is far better." □

Great Site Lines

Four more essential online baseball destinations

Baseball Primer www.baseballthinkfactory.org/files/primer
Links to significant baseball articles, accompanied by freewheeling, usually informed discussion threads.

Sons of Sam Horn pub208.ezboard.com/bsonsofsamhorn
Regular posters Larry Lucchino, Theo Epstein and Curt Schilling make this comprehensive Red Sox site a must-read.

Aaron's Baseball Blog www.aarongleeman.com
Twins-centric blogger Aaron Gleeman was worshipping at the altars of Justin Morneau and Johan Santana last season.

A Pinstriped Bible yesnetwork.com/yankees/news.asp?news_id=393
Going to YES for Yankees news is like reading *Pravda*, but these erudite, often hilarious weekly essays shine with objectivity.