

Version 1 of S

Here are some miscellaneous reflections and trivia. Much of this is drawn from handwritten notes and scraps of output in my files.

Dating the Conception and Birth

A group of us (basically the [five](#) mentioned before) began meeting in April, 1976. A memo of mine dated April 20 outlines the "the process of data analysis"; some notes from a presentation by Doug Dunn on April 21 go over software for time-series analysis; an undated note by Graham Wilkinson reviews data structures in GenStat. The next record is a note dated April 28, 1976 written by Doug Dunn, summarizing a meeting of the ``statistical computing group". This outlines some general points about future development of ``the SYSTEM" (its name would come [later](#)).

While interpreting these informal notes from so far in the past is hazardous, they seem to suggest that at this point a coherent view had not coalesced. There is a lot of terminology that in hindsight seems to dance around the idea of a single coherent language and system. For example, the concept of a ``general statistical software algorithm" is used for roughly what would become an S function, and the concept of an interface language has not yet surfaced (at least in the notes).

The next meeting was May 5, at which the notion of interfaces was presented as described below. I think this is probably the best date for the initial ``conception" of S. While the notes of April 28 leave us mostly in the environment of individual commands (in the style of GenStat and other packages current at the time) communicating with Fortran algorithms, my prepared notes for May 5, and some other material dated in the weeks after that, show more of the flavor of a language (a *programming* language, though we didn't say this explicitly) in which access to the underlying algorithms was made available.

Interfaces At the Very Beginning

The concept of an interface seems to be a recurring thread through the entire history of S. From my personal presentation at one of the first meetings (May 5, 1976) came a proposed structure for incorporating what we called algorithms, meaning Fortran subroutines, into the interactive language.

A hand-drawn diagram similar to the image on the left was used to illustrate the basic idea: an *interface routine*, here **XABC**, is generated to interpret the call from the interactive language as a call to the Fortran algorithm. The interface routine in Version 1 (and Version 2) was itself a Fortran routine, but programmed in an *interface language* that was then pre-processed into Fortran.

The interface language as initially proposed provided macros to define the arguments to the function in the interactive language and to generate other data to be used as temporary storage or for the result returned by the interactive function. These macros include ``type" declarations (initially only what we would later call basic vector types plus time-series and multi-way array).

All this was done, remember, by pre-processing into Fortran code that carried out the allocations. At some point, each interface routine was expected to then call a Fortran algorithm to do the real computing. This model applied not only to numeric computations but to graphics as well. The graphics capabilities, in fact, were a crucial part of the initial plan.

The First Few Months

Over the summer of 1976, some actual implementation began. The paper record has a gap over this period (maybe we were too busy coding to write things down). My recollection is that by early autumn, a language was available for local use on the Honeywell system in use at Murray Hill. Certainly by early 1977 there was software and a first version of a user's manual.

High on my personal FAQ list is ``Why is S called S?". In fact, we went along for some time without naming the system. I have an undated advertisement for ``An Interactive Language for Data Analysis" over the names of the five implementers of Version 1 (Becker, Chambers, Dunn, plus Jean McRae and Judy Schilling). This describes the language as ``being developed"; my guess is that it is from late spring or early summer of 1976. No name is offered for the language. The language is not explicitly offered to users in this note.

Rick Becker's [history](#) (page 5) describes the naming process, and dates it to July. We cast around for various names, getting suggestions from our colleagues, but without coming up with any satisfactory name that was not already taken. In the end, and with the C language as a precedent, we noted that the letter S was common to all the proposed acronyms, and chose the intersection. For the first couple of years, the letter was contained in single quotes, `S', but later undressed to a single character. (As we later found out, this had the side effect that we could not copyright it; single letters could not be copyright.)

[John Chambers<jmc@research.bell-labs.com>](mailto:jmc@research.bell-labs.com)

Last modified: Tue Mar 7 10:49:54 EST 2000