

THE STATUS OF FORTRAN 8x

Michael Metcalf

CERN, Geneva, Switzerland

ABSTRACT

Since 1979, the Fortran standardization committee, X3J3, has been labouring over a draft for the next version of the standard. Its initial intention of publishing this draft in 1982 was hopelessly optimistic, and in fact it was published only in 1987. A number of fundamental issues have been thrown up over the past three years, such that it is now clear that unanimity can never be achieved: efficiency versus functionality; safety versus obsolescence; small and simple or big and powerful? This paper reviews the current state and content of Fortran 8x, and attempts to bring out some of the controversial issues surrounding its development.

*FORTRAN is not a flower, but a weed. It is hardy, occasionally blooms,
and grows in every computer.*

A. Perlis

1. LANGUAGE EVOLUTION

The hardiness of the Fortran 'weed' is now an accepted, if remarkable, aspect of scientific and numerical data processing. Various reasons, or excuses, for its vigour have been put forward—the magnitude of the existing investment, its ease of use, and the efficiency of its implementations—but one which is sometimes forgotten is the fact that the language itself has evolved since it was first introduced in 1957. This oversight is most evident in the writings of its detractors, who often unfairly compare their own favourite language with code written in abysmal style in Fortran 66, conveniently ignoring the success of Fortran 77 which, given some self-imposed discipline [1], can be employed to write programs of a very high standard.

This evolution has its formal side. The language has twice been standardized in the framework of ANSI and ISO, in 1966 and 1978. Following the publication of the second standard, the technical committee responsible for the work, X3J3, re-formed to begin work on a third standard, with the intention of having it ready in 1982. This time scale was hopelessly optimistic, the draft being published only in 1987.

What are the justifications for continuing to revise the definition of the Fortran language? One is to modernize it in response to the developments in language design which have been exploited in new languages, particularly PASCAL and ADA. Here X3J3 has the advantage of hindsight, and can avoid the pitfalls of adding trendy features like the already outmoded DO...WHILE, whilst drawing on the obvious benefits of concepts like data hiding. In the same vein is the need to eliminate the dangers of storage association, to abolish the rigidity of the outdated source form, and to improve further on the regularity and portability of the language.

However, taken together, these would constitute only a ragbag of items perhaps better obtained by switching to other languages. The real strength of the new standard will be its incorporation of powerful array processing features and of derived-data types, allowing users to access vector processor hardware using a convenient notation, and to define and manipulate objects of their own design. Unfortunately for those of us who would like to see a new standard published by the end of

this decade, there are still many possibly insuperable obstacles to be overcome. Formally, X3J3 is currently at milestone eight of the eighteen milestones which have to be passed before a new standard receives full international recognition, and there are powerful voices stating that the advances it has made have gone too far, and that there should be some further retrenchment before it is acceptable to them. Such retrenchment would, of course, be unacceptable to others. A ballot of X3J3 held in April 1986, whose result was 16 to 19 against the draft of that date, resulted in a first slimming down, including the loss of BIT data type, in an attempt to reach a compromise between radicals who want a large range of new features and conservatives with more modest goals and who emphasize the importance of long-term backwards compatibility. A second ballot, 29 to 7, showed that better agreement had been reached, but that no true consensus had been achieved, given that two of the dissenting votes are those of IBM and DEC claiming to act as representatives of their users.

This dissension within the committee could have been regarded as healthy as long as it led to a constructive debate on its objectives resulting in a final resolution of the difficulties. However, some of the issues are so difficult, that reconciliation is now seen as unobtainable:

- Should Fortran 8x be innovative, or merely standardize existing practice?
- Should the language be small and simple, or big and powerful? Is the present proposal too big, making it impossible to fit into the small machines of 1990, impossible to be implemented by small software houses, and impossible to be understood by non-professionals (the bulk of Fortran' users)?
- Are users prepared to drop existing features, given 10 to 20 years' notice, or must all existing code work for ever?
- Are subsets of the language useful, or an impediment to portability?
- Do users want a safe language, or one which permits them to write the tricky programs more often associated with assembly languages?
- Are the existing proposals difficult and inefficient to implement? Does that matter if users thereby have an easier life?
- Will the heap storage mechanism necessary for dynamic array allocation impair efficiency, as happened with PL/1?
- Will the presence of new features cause existing ones to be implemented less efficiently?

In spite of all these points of disagreement, X3J3 and its parent committee X3 finally decided in the summer of 1987 that the draft of that time should be submitted to a formal period of public comment, at both the international and American levels. This period ended in February 1988, and showed that the public was on the whole unreceptive to many of the novel ideas contained in the document. By a margin of about two to one (interpretations vary) the 400 letters received seemed to want a simpler, less innovative standard, truer to the 'spirit of Fortran'. X3J3 has examined various proposals to simplify the language as a response to this negative reaction but, hamstrung by a two-thirds voting rule, is currently unable to agree on any single proposal. It might require direction from ISO to resolve this impasse, which could conceivably result in the work on Fortran actually being removed from ANSI and hence X3J3, and being taken over by another body*).

The development of the published document has not followed a straight path, and some items (for instance BIT data type and the significance of blanks) have been in and out several times. This has meant that the public has had some difficulty in following the committee's work. This present paper is an attempt to give a feel for Fortran 8x; fuller details are given in [2]. It reflects one *possible* outcome of the present debate, but without describing the form in which pointers are likely to be added.

*) At its meeting in September 1988, the ISO body (JTC1/SC22/WG5) defined its own version of Fortran 8x, which it renamed Fortran 88, and gave X3J3 five months to produce that as a new draft standard, or else!

2. BACKWARDS COMPATIBILITY

The procedures under which X3J3 works require that a period of notice be given before any existing feature is removed from the language. This means, in practice, a minimum period of one revision cycle, which for Fortran means a decade or so. The need to remove features is evident: if the only action of the Standards Committee is to add new features, the language will become grotesquely large, with many overlapping and redundant items. The solution likely to be adopted by X3J3 is to publish as an Appendix to the standard a set of two lists showing which items have been removed or are candidates for eventual removal.

The first list contains the Deleted Features, those which in the previous standard were listed as Obsolescent Features, and have now been removed. The list of Deleted Features is empty for Fortran 8x which contains the whole of Fortran 77.

The second list contains the Obsolescent Features, those considered to be redundant and little used, and which should be removed in the next revision (although that is not binding on a future committee). The Obsolescent Features are

Arithmetic-IF	Real and double precision DO-variables
Shared DO termination	DO termination not on CONTINUE (or ENDDO)
Alternate RETURN	Branch to ENDIF from outside block
PAUSE	ASSIGN and assigned GOTO

3. MAIN NEW FEATURES OF FORTRAN 8x

3.1 Source form

The new source form allows free form source input, without regard to columns. Comments may be in-line, following an exclamation mark(!), and lines which are to be continued bear a trailing ampersand (&). The character set is extended to include the full ASCII set, including lower-case letters (which in Fortran syntax are interpreted as upper case). The underscore character is accepted as part of a name, which may contain up to 31 characters. Thus, one may write a snippet of code such as

```
SUBROUTINE CROSS_PRODUCT (x, y, z) ! Z = X*Y
:
z1 = x(2) * y(3) -           &
      x(3) * y(2)
```

3.2 Alternative style of relational operators

In response to user demand, the deliberate redundancy of an alternative set of relational operators is introduced. They are

< for .LT.	> for .GT.
< = for .LE.	> = for .GE.
= = for .EQ.	< > for .NE.

enabling statements such as

```
IF (x < y .AND. z**2 > = radius_squared) THEN
```

to be written.

3.3 Specification of variables

The existing form of type declaration, as shown by

```
REAL a(5,5), b(5,5)
```

is extended to allow all the attributes of the variables concerned to be declared in a single statement. For instance, the statement

```
REAL, ARRAY(25), PARAMETER :: a = [25*0.], b = [25*1.]
```

declares the two objects *a* and *b* to have the attributes of being arrays of named constants, whose values are specified by the array constructors (see Subsection 3.9) following the equals signs. Many other attributes may also be specified, in particular real and complex variables may be specified to have a defined minimum precision and/or exponent range, as shown in

```
COMPLEX(PRECISION = 12, EXPONENT_RANGE = 100), ARRAY(10) &  
:: current
```

where the variable *current* is specified to have a minimum precision of 12 decimal digits and an exponent range of at least $10^{\pm 100}$. This facility is of great benefit in writing portable numerical software. Where constants corresponding to a given range and precision are required, a corresponding exponent letter must be chosen, for example

```
EXPONENT LETTER (12,100) C  
:  
current (3) = 15.6C57
```

At the time of writing, it seems likely that this facility will be simplified to correspond more closely to the actual, underlying hardware precision.

For those who wish to define explicitly the type of all variables, the statement

```
IMPLICIT NONE
```

turns off the usual implicit typing rules.

3.4 CASE construct

The CASE construct allows the execution of one block of code, selected from several, depending on the value of an integer, logical, or character expression. An example is

```
SELECT CASE(3*i-j)  
  CASE(0)           ! for 0  
  :                 ! executable code  
  CASE(2,4:8)      ! for 2, 4 to 8  
  :                 ! executable code  
  CASE DEFAULT     ! for all other values  
  :                 ! executable code  
END SELECT
```

The default clause is optional; one of the clauses *must* be executed.

3.5 Loop construct

A new loop construct is introduced whose syntax is combined with that of the old form of the DO-loop. In a simplified form it is

```
[name:] DO [(control)]  
    block of statements  
END DO [name]
```

(where square brackets indicate optional items). The control parameter, if omitted, implies an endless loop; if present, it may have one of two forms:

```
i = intexp1, intexp2 [, intexp3]
```

or

```
intexp4 TIMES
```

The optional name may be used in conjunction with CYCLE and EXIT statements to specify which loop in a set of nested loops is to begin a new iteration or which is to be terminated, respectively.

3.6 Program units

An enhanced form of the call to a procedure allows *keyword* and *optional* arguments, with *intent* attributes. For instance, a subroutine beginning

```
SUBROUTINE solve (a, b, n)  
    OPTIONAL, INTENT (IN) :: b
```

might be called as

```
CALL solve (n = i, a = x)
```

where two arguments are specified in keyword (rather than positional) form, and the third, which may not be redefined within the scope of the subroutine, is not given in this call. The mechanism underlying this form of call requires an *interface block* with the relevant argument information to be specified.

Procedures may be specified to be recursive, as in

```
RECURSIVE FUNCTION factorial(x)
```

The old form of the statement function is generalized to an *internal procedure* which allows more than one statement of code, permits variables to be shared with the host procedure, and contains a mechanism for overloading operators and assignment for derived-data types. This we shall return to in Subsection 3.14.

3.7 Extensions to CHARACTER data type

A number of extensions to the existing CHARACTER data type permit the use of strings of zero length:

```
a = ''
```

and the assignment of overlapping substrings:

```
a(:5) = a(3:7)
```

and introduce new intrinsic functions, such as TRIM, to remove the trailing blanks in a string. Some intrinsics, including INDEX, may operate, optionally, in the reverse sense.

3.8 Input/Output

The work in the area of input/output has been mainly on extensions to support the new data types, an increase in the number of attributes which an OPEN statement may specify, for instance to position a file or to specify the actions allowed on it, and to add a NAMELIST feature, illustrated by

```
NAMELIST/list/a, i, x  
READ(unit, NML = list)
```

which would expect an input record of the form

```
&LIST X = 4.3, A = 1.E20, I = -4
```

3.9 Array processing

The introduction of array processing features is one of the most important new aspects of the language. The reasons are threefold: an array notation defined in the language simplifies the syntax; new features extend the power of the language to manipulate arrays; and the concise syntax makes the presence of array processing obvious to compilers which, especially on vector processors, are able to optimize object code better.

An array is defined to have a *shape* given by its number of dimensions, or *rank*, and the *extent* of each one. Two arrays are conformable if they have the same shape. The operations, assignments and intrinsic functions are extended to apply to whole arrays on an element-by-element basis, provided that when more than one array is involved they are all conformable. When one of the variables involved is a scalar rather than an array, its value is distributed as necessary. Thus we may write

```
REAL, ARRAY(5, 20) :: x, y  
REAL, ARRAY(-2:2, 20) :: z  
:  
z = 4.0 * y * SQRT(x)
```

In this example we may wish to include a protection against an attempt to extract a negative square root. This facility is provided by the WHERE construct:

```

WHERE (x >= 0.)
    z = 4.0 * y * SQRT(x)
ELSEWHERE
    z = 0.
END WHERE

```

which tests x on an element-by-element basis.

A means is provided to select sections through arrays. Such sections are themselves array-valued objects, and may thus be used wherever an array may be used, in particular as an actual argument in a procedure call. Array sections are selected using a triplet notation. For an array

```

REAL, ARRAY(-4:0, 7) :: a

```

$a(-3,:)$ selects the whole of the second row, and $a(0:-4:-2, 1:7:2)$ selects in reverse order every second element of every second column.

Just as variables may be array-valued, so may constants. It is possible to define a rank-one array-valued constant as in

```

[1, 1, 2, 3, 5, 8]

```

and to reshape it to any desired form:

```

REAL, ARRAY(2, 3) :: a
a = RESHAPE([2, 3], [1 : 6])

```

where the first argument to the intrinsic function defines the shape of the result, and the second defines an array of the first six natural numbers.

3.10 Dynamic storage

Fortran 8x provides four separate mechanisms for accessing storage dynamically. The first is unlikely to survive the present carnage, and is not described here.

The second mechanism is via the `ALLOCATE` and `DEALLOCATE` statements which, as their names imply, are used to obtain and return the actual storage required for an array whose type, rank, name, and allocatable attribute have been previously declared in the procedure:

```

REAL, ARRAY(:), ALLOCATABLE :: x
:
ALLOCATE(x(n, m)) ! n and m are integer expressions
:
x(j) = q
CALL sub(x)
:
DEALLOCATE (x)

```

Deallocation occurs by default, unless the array has the `SAVE` attribute, whenever a `RETURN` or `END` statement in the same procedure is executed. The fact that allocation and deallocation occur in random order implies an underlying heap storage mechanism.

The third mechanism, useful for local arrays with variable dimensions, is the *automatic* array:

```
SUBROUTINE sub(i, j, k)
  REAL, ARRAY(i, j, k) :: x ! bounds from dummy arguments
```

whose actual storage space is provided (on a stack) when the procedure is called.

Finally, we have the *assumed-shape* array, whose storage is defined in a calling procedure, and for which only a type, rank, and name are supplied:

```
SUBROUTINE sub(a)
  REAL, ARRAY(:, :, :) :: a
```

Various enquiry functions may be used to determine the actual bounds of the array:

```
DO (i = LBOUND (a, 1), UBOUND (a,1))
  DO (j = LBOUND (a, 2), UBOUND (a, 2))
    DO (k = LBOUND (a, 3), UBOUND (a, 3))
```

where LBOUND and UBOUND give the lower and upper bounds of a specified dimension, respectively.

3.11 Intrinsic procedures

Fortran 8x defines about 100 intrinsic procedures. Many of these are intended for use in conjunction with arrays for the purposes of reduction (e.g. SUM), inquiry (e.g. RANK), construction (e.g. SPREAD), manipulation (e.g. TRANSPOSE), and location (e.g. MAXLOC). Others allow the attributes of the working environment to be determined (e.g. the smallest and largest positive real and integer values), and access to the system and real-time clocks is provided. A random number subroutine provides a portable interface to a machine-dependent sequence, and a transfer function allows the contents of a defined area of physical storage to be transferred to another area without type conversion occurring.

The MIL-STD 1753 bit intrinsic procedures are likely to be included in a final version [3].

3.12 Derived-data types

Fortran has hitherto lacked the possibility of building user-defined data types. This will be possible in Fortran 8x, using a syntax illustrated by the example

```
TYPE staff_member
  CHARACTER(LEN = 20)::first_name, last_name
  INTEGER::id, department
END TYPE
```

which defines a structure which may be used to describe an employee in a company. An aggregate can be defined as

```
TYPE(staff_member), ARRAY(1000)::staff
```

defining 1000 such structures to represent the whole staff. Individual staff members may be

referenced as, for example, `staff(no)`, and a given field of a structure as `staff(no)%first_name`, for the first name of a particular staff member. More elaborate data types may be constructed using the ability to nest definitions as in

```

TYPE company
  CHARACTER(LEN = 20)::name
  TYPE(staff_member), ARRAY(1000)::staff
END TYPE
:
TYPE(company), ARRAY(20)::companies

```

to build a structure to define companies.

3.13 Data abstraction

It is possible to define a derived-data type, and operations on that data type may be defined in an internal procedure. These two features may be combined into a module which can be propagated through a whole program to provide a new level of data abstraction. As an example we may take an extension to Fortran's intrinsic `CHARACTER` data type whose definition is of a fixed and pre-determined length. A user-defined derived-data type, on the other hand, may define a set of modules to provide the functionality of a variable length character type, which we shall call *string*. The module for the type definition might be

```

MODULE string_type
  TYPE string(maxlen)
    INTEGER::length
    CHARACTER(LEN = maxlen)::string_data
  END TYPE String
END MODULE String_type

```

With

```

USE string_type
:
TYPE(string(60)), ARRAY(10)::cord

```

we define an array of 10 elements of maximum length 60. An actual element can be set by

```
cord(3) = 'ABCD'
```

but this implies a re-definition, or *overloading*, of the assignment operator to define correctly both fields of the element. This can be achieved by the internal procedure

```

SUBROUTINE c_to_s_assign(s,c) ASSIGNMENT
  TYPE (string(*))  ::s
  CHARACTER(LEN = *)::c
  s%string_data = c
  s%length      = LEN(c)
END SUBROUTINE c_to_s_assign

```

which can be included in the module, together with other valid functions such as concatenation, length extraction, etc., to allow the user-defined string-data type to be imported into any program unit where it may be required, in a uniformly consistent fashion.

4. CONCLUSION

This paper has tried to set out the background to the current revision of the Fortran language standard, and to give an overview of its principal new features.

Any standard is a compromise between conflicting points of view, as indicated in the first section. On the other hand, the presence of a widely accepted and implemented standard provides the only practical means to write and to use portable programs, and the evolution of that standard is the only means to escape from the temptation to use vendors' extensions or private dialects and languages. It is my opinion that although the current proposal does not contain everything which we would like to see, nevertheless, its final acceptance and implementation are the only way to avoid a future of anarchy and chaos. It is to be hoped that reason will prevail, and that the present serious disagreements can finally be resolved to the satisfaction of the largest possible section of the user community.

* * *

REFERENCES

- [1] M. Metcalf, *Effective FORTRAN 77* (Oxford Univ. Press, Oxford, 1985).
- [2] M. Metcalf and J. Reid, *Fortran 8x Explained* (Oxford Univ. Press, Oxford, 1987).
- [3] BITPAK, M441, CERN Program Library.