FAMULUS REDIVIDUS : A case history in software development

1. Famulus Past

1 “The real father of FAMULUS,” writes Theodore B. Yerke, “was Douglas C. Engelbart.” [1]. Engelbart was an eminent theoretician in that hotch-potch of behavioural science, cybernetics, psychology and librarianship for which, in the early sixties, the name “information science” had yet to be coined. His name was associated with the radical proposition that the information needs of the individual researcher were not necessarily identical with those of the librarian, and that information systems should be at least as well designed for the former as for the latter, since “It is the problem-oriented individual who is the basic module in our intellectual community. It is his effectiveness which must be the ultimate concern of anyone wishing to see that community perform better.”[2] In these Thatcherite days, that yoking of 'effectiveness' with 'intellectual' may be so startling that we overlook the implication, more remarkable in the historical context within which it was articulated, that the idiosyncratic response is of primary importance, that the progress of knowledge is not simply a matter of greater pooling of resources and ever more effective standardisation.

2 We would also do well to recall the type of facilities available in 1964 for the researcher wishing to use the new technology of the computer age for his private documentation needs. Communication with the giant 'electronic brains' of those days was exclusively performed through the mediation of decks of punched cards, usually prepared by specially-trained staff on a pre-printed proforma. Even by 1970, the use of interactive teletypes was still a novelty at most research installations. Bibliographical information, if available in computerised form at all, came in a format designed for and by librarians which could not easily be tailored to individual taste or priorities. In 1970, Yerke wrote, “It is not yet widely understood that personal information systems are necessary to researchers, because their orientation is basically antithetical to that of the formal system.”[3]. Small wonder then that researchers working close to the frontiers in their own field continued to rely on manual methods of organising that informal nexus of hearsay, gossip, purloined abstracts and quick notes in which the only really useful information is to be found.

3 In 1964, Yerke, librarian at the USDA Forestry Service's Pacific S.W. Experimental Station in Berkeley California, began to explore the possibilities of computerised personal documentation systems, apparently largely as a result of pressure from scientists at the station wishing to have their personal bibliographies maintained on the library's existing cataloguing software. This was a loosely organised collection of routines called CATALOG which had been written by the station's chief programmer, Robert M. Russell and ran on an IBM 7090 at UCLA. Yerke was encouraged by the demand to build upon the experience gained through using CATALOG and to that end he applied for funds from the central offices of the U.S. Forestry Service in order to develop a better integrated package with what we would now call a more user-friendly interface. For this new package, he chose the name FAMULUS, according to Webster's 7th dictionary the word used for a sorcerer's apprentice, although as the OED more prosaically points out it is simply borrowed from the Latin word for any household servant, and is cognate with both 'family' and 'familiar'. As a Germanist, Yerke presumably had Goethe's Faust and his Mephisto in mind. [10]

4 The intention was to develop a system which could be used throughout the Forestry Service and portability was therefore of more than usual importance. The programs would be written in the new ANSI standard Fortran66, with only a few routines hand-coded in Assembler for efficiency. Russell and Yerke had completed the design by the end of 1966 and Hilary D. Burton began coding

the new package in 1967, initially for the Forestry Service's own CDC 6400 machine, on which Famulus first ran in June 1968. An implementation for the UCLA's new IBM 360 followed shortly afterwards. [4,5]

5 The rapidity with which new versions appeared and the range of applications for which Famulus suddenly became the only solution form vivid testimony to prescience of the system's designers. To reconstruct a full stemma, charting the progress of the software, would be a task almost as complex and perhaps as futile as to list all the uses to which it has been put. [11] Within a few years of its original appearance, versions of Famulus were to be found across the American continent, being used by lawyers, doctors, historians and librarians as much as by researchers in forestry or other sciences. By 1972 it had reached New Zealand, Australia, India and Europe.

6 Famulus was introduced to the U.K. in 1971 by Francois Crompton-Roberts at the London School of Hygiene and Tropical Medicine. Together with Alan Shaw at University College, London, Crompton-Roberts made several extensions to the package, notably the addition of the KEY, KWIC, EDITP, COUNT and MULTIPLY programs.[6] These extensions, which were subsequently propagated by a host of UK university implementations, mostly reflected a bias towards more specifically textual and less straightforwardly bibliographic applications, corresponding to a shift away from the purely scientific in the typical Famulus user. For example, when the package was implemented on the ICL 1906A at the Science Research Council's Atlas Laboratory in 1976, extensions were made to support use of non-Roman alphabets [7]. Several other UK “third generation” implementations attempted to simplify the preparation of data for input to Famulus and to vary the format of its output.

7 In January 1981, an open meeting was convened by the Inter-University Software Committee of the Inter-University Computing Council (the independent body which represents UK academic computing centres to the Computer Board, the government body which is their paymaster) to discuss the future of the Famulus package. Although it was now well established, in some variant form or another, at most sites, no one site could be expected to take responsibility for support or development of software so variously re-implemented. Yet both support and development were necessary if use of the package was to continue perhaps for another decade. Perhaps, on the other hand, demand for it would abate as more complex and sophisticated database management systems became more widespread. As is customary for such meetings, a working party was

appointed to investigate and report back. Active membership of the Working Party, chaired by Dr A.J. Duke of Leeds University, naturally varied during the three years of its existence; its core members were Mrs E.M. Mandl of Manchester Regional Computing Centre, who acted as secretary, Dr J. Read of Edinburgh Regional Computing Centre, Mrs K. Crennell of the Rutherford Atlas Centre and the present author.

8 The Working Party's first task was simply to investigate the current state of play, to which end computing services in UK Universities, Polytechnics and research council funded sites were all surveyed during the summer of 1981. A questionnaire was sent to over 200 sites, asking which version of Famulus was in use (if any), setting out the Working Party's current plans for enhancing the package and soliciting respondents' views on these or other changes to the software. The responses, set out in a report to the IUSC in October 1982 [8] confirmed that the package was still widely used, at around 40 different sites, each supporting anything from two or three to several dozen applications. Despite the package's origins, it was apparent that well over half of all Famulus applications were not private documentation systems of the type Yerke had had in mind, but systems designed for several co-operative users. Although unable to quantify demand for the package, most sites seemed sure that it showed no sign of abating, while the soundness of the original design is perhaps confirmed by the fact that the enhancements most frequently requested were all largely cosmetic (better report facilities, easier input, clearer diagnostics etc) rather than radically different in approach. It was clear that effort put into an integrated and consistent version of the package re-written in standard Fortran77 would be of use to the academic community, and the Working Party therefore began this task in 1983.

9 At first, the Working Party hoped that a “quick respray job” (to use the Chairman's phrase) would be possible within a year. As design of the new software advanced, the optimism of this view became evident. By the summer of 1983, (when an earlier version of this paper was presented at an IUSC Workshop on Non-numerical Computing) over half the new package was under test and the bulk of the re-design was complete. Because none of the Working Party members was occupied full time on the project, meeting only at roughly monthly intervals when other commitments permitted, a somewhat unusual modus operandi evolved. Sections of the package were allocated to different members for writing which when generally available would then be tested in parallel on the four different machine ranges in use by the Working Party's members. This distributed style

of development coincided with both the development of the UK Joint Academic Network (JANET) which enabled code and messages to be distributed electronically and the availability of a new generation of Fortran compilers on most main frames, with the consequence that at times the Working Party seemed to be devoting more effort to exploring bugs in these new facilities than to developing software. It is perhaps no mean tribute to human nature that the members of the Working Party managed to remain on speaking terms not only with each other but also with local Fortran and communications support staff.

10 Despite these problems, work continued steadily throughout 1983. The first Famulus77 internal format file was created in April, and a substantial portion of the package was available for test by October 1983, when the Working Party reported back to the IUSC formally requesting that funds be sought from the Computer Board to complete the project. The first complete version was under test by June of 1984. A users manual, written by the present author, was also produced at this time. The requested funds having been obtained, Mr J. R. Martin was appointed to complete the rewriting of the package, test it and set up a mechanism for assisting in its trial distribution to a number of sites, for one year from October 1984. Shortly after Easter 1985, all Computer Board funded sites were circulated with the offer of a free trial version of the package, which several took up. As anticipated, there were very few problems in porting the software to machines as divers as the Harris super mini and the DEC 20. By the end of 1985, the new package was in use at forty different UK sites, with implementations planned or running for ICL 2900 (VME, EMAS), Honeywell (Multics), CDC Cyber (NOS/BE), Harris (Vulcan), Gould (Unix), Digital VAX (VMS) and DEC-20 (TOPS), IBM (VM/CMS, MVS, MTS), Prime (PRIMOS) and GEC 4090 (GCOS).

11 In January 1986, five years after the original IUSC open meeting, an agreement was concluded between the Famulus Working Party and the Program Library Unit (PLU) of Edinburgh University. PLU, which distributes several major software packages to UK Universities and elsewhere, agreed to take on the marketing of the new package (at very favourable rates for academic institutions) and to maintain it so far as is practicable. More significantly, in view of the package's chequered history, PLU also agreed to co-ordinate any future development of the Famulus77 package.

2. Famulus Present

12 Space and time both preclude a full discussion here of all the features of the new software. As these are in any case freely available in the new Users Manual [9], a short summary only is given here, highlighting differences between the old and the new packages.

13 Famulus is a simple text file system and not a true database management system. It is intended for use in areas such as cataloguing or bibliography where large amounts of very variably structured textual data are to be stored, maintained and retrieved. It is very simple to use, requiring no programming skill or experience but is not primarily an interactive system, although it can be run from a terminal, using a simple English-like command language.

14 Famulus was originally conceived as a suite of nine independent batch programs, each performing a distinct operation. Program Edit, for example, added or amended citations to a file; Galley printed them out; Sort re-ordered them to produce a new file, and so on. Selecting citations according to their content was also performed by a distinct program, athough selection by citation number was available from most programs. This had two drawbacks: firstly any reasonably complex operation (e.g. add citations, sort the result, select and print a subset of the result) required several program runs. In the days when job control was performed by decks of cards, most users probably did not see the resulting card-shuffling as particularly onerous (indeed, it was probably delegated); now that job control is more usually performed by typing at a keyboard, some sort of easy chaining of commands seems essential. A second consequence of this lack of integration was that new components also had been added piecemeal, so that facilities which the user might reasonably expect to be available throughout the package were in fact available only in certain programs - the ability to operate on a subset of fields, for example. Thirdly, the overall size of the program had become unnecessarily swollen by considerable duplication of code with common functionality.

15 To aid transition for users of the old Famulus, it is still possible to operate one step at a time; several Famulus programs may be chained together, each step using as input the output from the previous one. However, because most functions are now available from most programs, there is less need to do this.

16 A similar type of integration was attempted in the command language of the package. Again the aim was to preserve the basic form as far as possible. Commands can be given in any order, subject only to the restriction, common to both old and new versions, that when a program uses an existing

file as input, the command nominating that file (ID in the old system, INPUT in the new) must be given first. The basic syntax of the command language has not been changed, to retain the familiarity which the package has long enjoyed, but some of its more irritating restrictions and accreted eccentricities have been removed. Commands may be given in upper or lower case or a mixture, and the special characters used to delimit parts of the language (commas etc.) can be re-defined by the user.

17 A Famulus "database" is a single file consisting of many citations, each of the same structure. This structure is defined when the file is first created and consists of a list of fields, each of a defined name and type. Types available are integer, date and string. There is no limit on the size of any string field, other than the over all limit on the size of a citation. Integer and date fields are displayed in the format in which they are entered, but sorted and compared in a normalised form to ensure correct comparison. Up to fifty different fields may be defined for a file; not all of them need be present in each citation of course, and no space is taken up by data stored in unused fields. New fields cannot be defined for an existing file, but the Ossify program provides an easy way of dumping all or part of an existing database in a portable format which can then be reloaded into a new structure.

18 For upwards compatibility, the dump format produced by the new Ossify program is identical with that produced by the old. It is thus very simple to convert existing Famulus databases to the new system. The new Ossify program additionally supports both subsetting of the input file (by content or field type) and renaming of fields.

19 Preparation of citations in the external format remains the only way of loading data into a Famulus database. Utilities for interactive creation of such citation files were common extensions to the old Famulus, and the new system is no exception. However, provision of a fully integrated "front end" to the package remains unfinished business. Many users prefer in any case to us the editor or word processing package with which they are all ready familiar. The new Famulus does have extensive batch editing facilities which allow existing citations to be amended or expanded. This has a number of uses aside from the correction of errors. If for example, tagging has been used systematically to indicate such features as changes of typeface, it is a simple matter for Famulus to convert all such tags to the appropriate codes for driving a typesetter or other output device.

20 Citations in a file may be sorted according to the contents of one or more fields. A default sorting order is defined when the file is created, but may be changed at any time. Special characters, such as diacritics or padding characters, may be specified and particular characters may be equated to ensure that citations are sorted correctly. It is also possible to define “nonsorting” words, that is, particular words which are to be ignored for sorting purposes when they occur at the start of a field. The old Famulus had this facility “hard-wired-” the strings 'A' 'AN' and 'THE' were always disregarded. While excellent for English, this was less useful for other languages. In German, for example, 'AN' should not be so treated, but 'DIE' should. The new system allows the user to specify exactly which terms are to be so treated.

21 Selection of citations by their content is available for all programs in the new version, as mentioned above. A simplifed form of the usual type of Boolean expression is used, in which quite complex selection criteria can be rendered economically. Searches may be exact or fuzzy, the syntax indicating unambiguously whether the search is to find the string 'ABC' occurring as the whole of a field, as a single word in a field or as a substring of a word in a field; it is also possible to perform arithmetic comparison on the contents of date or integer fields (e.g. search for dates before 1900, values greater than 24 etc.) which was not possible in the old package. Some examples of the new syntax are given below:

22 SELECT IF: AUTH .EQ. “SMITH” selects any citation in which the AUTH field contains only the word SMITH

23 SELECT IF: KEYW .INC. “%OLOGY.” OR. “%ISM” selects any citation in which the KEYW field contains at least one word ending in either OLOGY or ISM.

24 (Note the use of the % sign (omnibus) to indicate an arbitrary number of characters; the actual character used may be redefined by the user using the OMNIBUS command.)

25 SELECT IF : AUTH .AND. TITLE .INC. “ROYAL WEDDING” selects any citation in which the phrase “ROYAL WEDDING” occurs in both AUTH and TITLE fields.

26 SELECT IF : KEYW .OR. TITL .INC. “POTATO.” AND..NOT. “POTATO BLIGHT” selects any citation in which the word POTATO occurs in either the KEYW or the TITL field, except for those in which it is followed immediately by the word BLIGHT.

27

28 For users who dislike the Fortran style relational operators, most implementations of the new package permit the use of single-character alternative symbols.

29 Reflecting widespread discontent with existing facilities, the new Famulus gives the user considerably more control over the appearance of selected citations when they are printed out. Editing commands such as REPLACE, FILL, PREFIX etc can be used to alter the contents of individual fields or all fields of a specified type. The size of the Report page and its margin may be re-defined. Citations may be presented in one of four formats TAGGED (formerly known as PRINT BY FIELDS), CATALOGUE (formerly PRINT BY SUBJECTS - an added refinement is that the number of levels of heading used is also user-definable), COLUMNS (in which citations are presented as a table of values) and COMPACT in which each citation is compressed if possible to a single record, perhaps for re-input to some other formatter. These facilities are available in every program which produces printed citations, i.e. every program except Index, Kwic and Ossify.

30 Index and Kwic produce specialised reports which give Famulus its text processing (as distinct from bibliographic) power. In conjunction with the well known Key and Multiply options of the old Famulus, these enable the researcher to investigate the vocabulary of a set of fields, to re-index arbitrarily and ultimately to construct his own thesaurus of terminology. Terms to be indexed are delimited by user-specifiable punctuation and may be taken from one or more nominated fields. An inclusion or an exclusion list can be specified, as may a list of synonyms; any of these lists may include omnibus characters, so that (for example) a single synonym may be specified for all terms begining or ending with a specified set of characters. References in the index produced may use the contents of a nominated field to identify the citation in which the term occurs, an extension of particular usefulness when citations include some identifying attribute such as an accession number or catalogue reference.

3. Famulus Future

31 It is a little early to tell whether the new Famulus will be as successful as the old, but early indications are promising. At Oxford, it is being used to prepare bibliographies for publication, using the batch editing facilities mentioned above to insert LASERCOMP typesetting codes. At the University of Nottingham, it is being extensively used in the analysis of mediaeval parish records. In Manchester, it is planned to move a large Museum database to the new system.[11]

32 In retrospect, the causes of Famulus's continued popularity are probably its basic simplicity and the shallowness of the learning curve involved in getting the best out of it. For users brought up on traditional manual bibliographic or documentation systems, Famulus presents very few problems, once the initial keyboard phobia has been overcome. As a text manipulating system, even today, there are very few commercially available sophisticated database management systems to emulate it. Its attractiveness to the individual researcher probably is as much to do with the facilities it offers for him or her to define his or her own idiosyncratic structure without needing to learn any jargon. Although its facilities are therefore limited, they coincide with those of many users. And finally, the software has always been (and remains) very cheap, very portable and almost ubiquitous.

33 Paradoxically, the goal of Engelbart and Yerke has still not been reached by Famulus. Over half of its current user population seems to be drawn from research groups rather than individuals. Perhaps, just as Famulus was ahead of its time in the user-friendliness of its software interfaces, so it was also ahead of its time in the type of hardware for which it was best suited. Now that personal computers are available of a power adequate to perform the sort of tasks at which Famulus excells, we can say that the hardware has at last caught up with the original concept of the personal documentation system. Famulus77, when it is available on the new generation of micro computer, will perhaps still be with us in another two decades.

NOTES-LABEL

1 [1] Theodor B. Yerke, private communication, 1982

2 [2] Douglas C. Engelbart. Special considerations of the individual as a user, generator and retriever of information. American Documentation 12 (1961), pp 121-25.

3 [3] Theodor B. Yerke. Computer support of the researcher's own documentation. Datamation 16 (1970) p.75.

4 [4] Theodor B. Yerke. Das FAMULUS - Computerprogramm als Hilfe zur individuellen Befriedigung des Informationsbedarfs von Naturwissenschaftler. Mitteilungen der Gesellschaft fur Bibliothekswesen und Dokumentation des Landbaues 18 (1972) pp 97-115

5 [5] Hilary D. Burton; Robert M. Russell; Theodor B. Yerke. FAMULUS: a computer-based system for augmenting personal documentation efforts. American Soc Info Sciences, Proceedings 6 (1969) pp 53-56

6 [6] Alan Shaw. FAMULUS Reference Manual. University College London Computer Centre, 1974.

7 [7] Susan M. Hockey. FAMULUS on the 1906A. Atlas Computer Division, Rutherford Laboratory, 1974.

8 [8] IUSC Famulus Working Party, Interim Report, October 1982

9 [9] [Burnard, L.D.] Famulus77 Users Manual. Oxford University Computing Service, 1985.

10 [10] I am indebted to Francois Crompton-Roberts for this observation.

11 [11] A bibliography of published references to Famulus and work done with it is in preparation; copies are available from the author, who would also much appreciate pointers to any ommissions from it.

ABSTRACT-LABEL

This paper describes the origins and history of Famulus, a free text retrieval system originally designed twenty years ago to support personal research bibliographies, but much extended and re-implemented. A new integrated version of the package consolidating most of the enhancements made in the past has been developed by a consortium of UK Universities and is now being marketed: this paper describes its facilities in the context of the original package design.

AUTHOR-LABEL