

EUROMATH

Task 4.3: Requirements Analysis

Requirements Analysis Report

Kjeld Schmidt
Gordon McAlpine

Dansk Datamatik Center

Classification:	OPEN, publicly available
Document No.:	EUROMATH/DDC08/1988-10-13
Copyright:	1988, Dansk Datamatik Center

Table of Contents

Preface.....	3
1. Characteristics of the Mathematical Community.....	4
2. Scientific Discourse	6
2.1. Private Discourse.....	6
2.2. Public Discourse	7
2.2.1. Research Area Networks.....	7
2.2.2. Preprint Distribution	8
2.2.3. Electronic Publishing	13
2.2.4. Bibliographic Information Retrieval.....	14
2.3. The Euromath Directory.....	16
2.4. Personal Databases.....	16
2.5. The Euromath Information Delivery and Retrieval System	16
3. Document Preparation.....	19
3.1. Interface Issues.....	19
3.1.1. Direct Manipulation	19
3.1.2. WYSIWYG.....	19
3.1.3. Interaction Style	20
3.2. Alphabets and Fonts.....	20
3.3. Document Structure.....	20
3.4. Outlining	21
3.5. Illustrations	21
3.6. Other Facilities.....	22
4. Research Support Tools	23
4.1. Symbolic Manipulation.....	23
4.2. Graphics.....	23
5. Integration.....	24
5.1. Different Types of Integration.....	24
5.1.1. Inter-Workstation Integration.....	24
5.1.2. Application Integration.....	24
5.1.3. User Interface Integration	25
5.2. Integration vs. Flexibility	25
5.2.1. Document Processing.....	29
5.2.2. Symbolic Manipulation and Document Processing.....	30
5.2.3. Information Delivery and Retrieval.....	31
5.3. A Feasible Euromath Architecture.....	31
Literature.....	33

Preface

Design of large-scale computer systems for complex organizational environments all too often results in systems that can only be characterized as failures, not in the technical sense (bug ridden software etc.), but in the sense that they do not fit the form of cooperation that characterized the particular organizational setting. At best, the misfits are not accepted by the putative target group; at worst they disrupt cooperative efforts instead of enhancing them.

Whereas a Darwinistic approach to software development (a spontaneous trial-and-error process) may be feasible in the design of minor single-user applications, it is not feasible and may have catastrophic consequences in the design of large-scale systems supporting cooperative work. In the latter case, a systematic analysis of the domain in question is required.

Thus, in order to ensure that the EUROMATH system will meet the needs and requirements of the mathematical community, a systematic requirements analysis has been conducted. The analysis has been based upon interviews with approximately 20 mathematicians as well as the findings of various contributions to the sociology, psychology, history and and philosophy of mathematical research.

This paper contains a presentation of the findings and conclusions of the Requirements Analysis. The “Requirements Analysis Report” explains the rationale for the requirements for the EUROMATH System by describing the results of the requirements analysis, whereas the purpose of the “Requirements Specification”¹ is to concisely define the requirements for the EUROMATH System.

The National Institute for Higher Education (NIHE), Dublin, has contributed to the analysis of the current system of publication and preprint distribution.

¹ G. McAlpine and K. Schmidt: “Requirements Specification”, EUROMATH/DDC09/1988-10-13.

1. Characteristics of the Mathematical Community

“Like the hordes and horses of some fabulous khan, today’s mathematicians have ridden off in all directions at once, conquering faster than they can send messages home.”

David Bergamini²

The general characteristics of the organization of mathematical research have important consequences for the requirements to be met by the EUROMATH system.

(1) Compared to other scientific disciplines, the problem domain of mathematics appears strikingly ‘flat’. As pointed out by Charles Fisher:

“In contemporary physics, the investigation of elementary particles is generally acknowledged to be the most fundamental area of research. Many physicists feel that theory is forged in attempts to understand the interactions of particles and that theory is then passed down to other areas of physics. Hence, there is a generally accepted notion of what is fundamental and a vaguely defined hierarchy of prestige based on theory production. Mathematics, on the other hand, is a more diffuse, less focused discipline, the members of which often work in relative isolation.”³

Thus a mathematician working in one speciality need not have a feeling that others are working on more fundamental problems than he himself.⁴

Accordingly, theoretical changes in mathematics do not propagate in the top-down manner of physics. As observed by Richard Startup:

“There is, in fact, evidence that change in mathematics takes a more diffuse form than in (say) physics. Mathematics appears more highly specialized and no one set of problems is universally acknowledged as basic. The tendency is for theoretical or methodological innovation to lead to a reorganization of specialist areas. Theories do not replace theories in the sense that Einsteinian replaced Newtonian mechanics: new theories may, rather, solve some of the problems generated by existing theories without disproving them. Many of the problems - including those raised by relationships to other fields - may remain undisturbed.”⁵

(2) Mathematics is characterized by being conducted in highly specialized research areas.⁶ Typically, the number of mathematicians engaged in a given research area is relatively small. The approximately 50,000 mathematicians in the world are distributed on more than 4,500 specialties in the mathematical “Subject Classification” scheme compiled by *Mathematical Review* and *Zentralblatt für Mathematik* in 1980, that is 11 mathematicians per specialized research area. We may assume, of course, that a mathematician is active in more than one field. If we assume that an average mathematician is active in three fields - and on the basis of interviews with numerous mathematicians this seems to be a realistic figure - we arrive at an average number of mathematicians per research area of 33.

² Quoted by Hagstrom: *The Scientific Community* (1965), p. 228.

³ C. S. Fisher: “Some Social Characteristics of Mathematicians and Their Work” (1973), p. 1098.

⁴ C. S. Fisher: “Some Social Characteristics of Mathematicians and Their Work” (1973), p. 1099.

⁵ R. Startup: “The Sociology of Mathematics” (1980), p. 157.

⁶ Cf. Hagstrom: *The Scientific Community* (1965), pp. 227 ff, and R. Startup: “The Sociology of Mathematics” (1980), p. 157.

Not surprisingly then, a mathematician who is well established in his field will tend to know most of the colleagues in the area, at least he or she will know of their work. This is the normal situation according to numerous interviews. For example, one of the informants observed:

“In reality, a typical mathematician is member of a community that is far less than the mathematical community at large.

If you count the number of colleagues a mathematician is in direct research related contact with, in the sense that they exchange papers, meet at conferences etc., then that number in some specialties is as low as 10 people globally.

Globally?

Yes, globally! That is, those who really are capable of assessing the details of what A is doing. (It depends on the subject, of course - some research areas are *en vogue*.) On top of that, there is a much wider network of people engaged in familiar work. So, if you are active in research, you are member of a family of, say, 200 people. And you don't know all of these 200 people, perhaps only 50 of them.”

Furthermore, a mathematician will have more or less regular correspondence with a large selection of his colleagues in the area. According to numerous statements from mathematicians, the number of colleagues with whom a mathematician has regular correspondence is typically 20-30. The fact that this number is of the same magnitude as the number of mathematicians per specialized research area as estimated above seems to confirm the subjective assessment of many mathematicians that they know most of the active mathematicians in their field.

(3) Due to the very high degree of specialization of mathematical research, the potential audience of a typical mathematical paper is very small. As noted by an interviewee:

“If you close your eyes and pick a paper from the 1600 mathematical journals you will obtain, with a very high degree of probability, a paper that no one ever reads, except for the reviewers.”

(4) As a rule, the colleagues with whom a given mathematician is corresponding are located in different countries and even continents. For example, of the internationally co-authored articles with Danish participation, in the period 1972-77, 46 pct. were written jointly with North American partners.⁷

⁷ Schøtt: *Fundamental Research in Small Countries. Mathematics in Denmark, 1928-1977* (1979), p. 108.

2. Scientific Discourse

As any other science, mathematical research is a highly cooperative endeavor. Thus, from the outset the primary objective of EUROMATH has been to enhance the ability of European mathematicians to cooperate.

Several different modes of conducting scientific discourse can be distinguished. Discourse may either be conducted privately, i.e. within a closed ensemble of researchers and in principle for the eyes of the members of the ensemble only, e.g., correspondence between colleagues, or it may be conducted publicly, i.e. in principle - albeit not in practice - accessible to any mathematician, e.g., publication of a paper or reviewing a paper.

2.1. Private Discourse

For normal research related correspondence between mathematicians, speed of delivery does not seem to be an issue. Most mathematicians seem to be fairly satisfied with ordinary postal services for normal research related correspondence. However, for various reasons, a more rapid means of correspondence is required:

(1) In some countries, ordinary postal services perform badly. For instance, a letter to Italy or Spain may take a month. That is not tolerable. Electronic mail can alleviate this problem.

(2) It is becoming fairly common for mathematicians to collaborate on research projects in the course of which they will write papers and books together. Thus about one third of all mathematical papers have more than one author. Some of these papers are, of course, produced by people working at the same location. But a large, and increasing, part of these papers are co-authored by mathematicians separated by wide distances. (Cf. fig. 1).

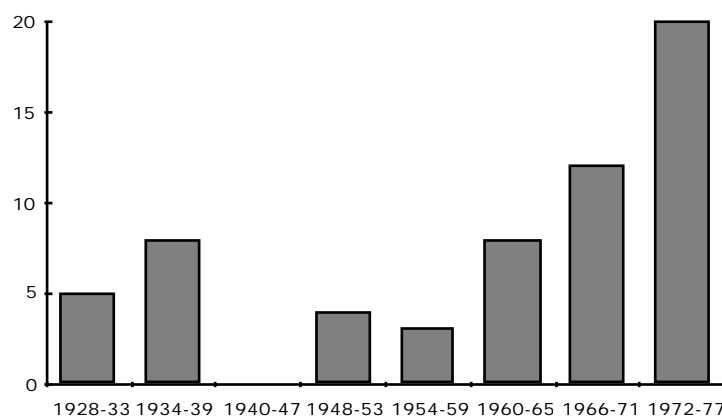


Fig. 1. Percentage of articles by Danish mathematicians in foreign languages that are internationally co-authored, by period, 1928-1977.⁸

⁸ Schøtt: *Fundamental Research in Small Countries. Mathematics in Denmark, 1928-1977* (1979), p. 82.

(3) In applied mathematics, formally organized cooperative projects are becoming common. This implies needs for frequent communication for purposes of setting up and running projects, that is, the same needs that can be observed in engineering and computer science.

In these situations speed of delivery is definitely an issue. Electronic mail can obviously alleviate this problem. By exchanging notes, comments and drafts electronically the collaborators can speed up communication immensely. This gain has its price, however. A hard copy of a draft allows the reader to enter comments in a very direct and succinct way with various intuitively understandable graphical symbols. This kind of facility is not available in current commercial software products; it is still a research issue. Also, speed of reading from current video display screens is significantly inferior (by about 30%) to reading from a sheet of paper, although a recent study indicates that this problem can be overcome.⁹ Because of these shortcomings, mathematicians engaged in collaborative writing and communicating by means of electronic mail typically send hard copies of draft versions so as to enable the reader to enter comments directly.

In general, mathematicians communicate sporadically. Not because of limitations in the communication facilities but because of the mode of working and cooperating that distinguishes mathematicians. This has implications for electronic mail facilities. It is likely that mathematicians will tend to restrict their use of electronic mail facilities to such situations where they are engaged in collective tasks (conducting a common research project, writing a paper or a book, sitting on a editorial board or a committee, etc.) In other words, mathematicians will be casual users of electronic mail facilities. The implication of this is that the commands of the electronic mail facility of the EUROMATH system should be very easy to learn and remember.

The sender of a message via electronic mail has no way of knowing in advance if and when the addressee will look for mail. In many of the above mentioned collective tasks the sender of a message needs to make sure that the recipient has read the message within some time limit. Thus, the electronic mail facility should provide a facility whereby a user can ascertain if his messages have been delivered and read.

The recipient of messages must be able to decide whether and when he or she wants to read incoming messages. Messages must have a 'header' providing information on the name of the sender, the date, the subject (keywords, possibly abstract), the size of the document, etc.¹⁰

2.2. Public Discourse

2.2.1. RESEARCH AREA NETWORKS

The fact that mathematical research is normally conducted within relatively small research networks makes mathematics an area of potentially successful comprehensive application of computer mediated conferencing.

⁹ J. D. Gould et al. "Why reading was slower from CRT displays than from paper", *Proc. ACM CHI+GI 87, Toronto*, pp. 7-11.

¹⁰ Cf. Lyne DaSylva: "First Field Study of Electronic Mail and Conferencing System", EUROMATH, Task 2.1.3.: Field Study of Interpersonal Communication, DDC, 1987.

Computer mediated conferences cannot easily be distinguished from conventional electronic mail. For instance, electronic mail may be distributed to a wide audience while access to computer mediated conferences may be restricted to a specific group of users. The difference is not of a technical nature but a difference in the mental model suggested by the system. Firstly, a conference system suggests the notion of a public memory, shared by the participants and not erasable, whereas an electronic mail system suggests a notion of transient and parochial relations of shared memory. A contribution to a conference virtually takes on its own life, independently of the ensuing actions of the participants, whereas a message conveyed by means of electronic mail ceases to exist unless the recipients take action to save it.

Computer mediated conferences may be moderated (all contributions are first routed to a specific user, called a moderator, who then makes sure that only relevant contributions are passed on to the audience of the conference). Moderated conferences covering as many mathematical research areas as possible should be organized and launched as part of the design and implementation of the EUROMATH system. It would be ideal if reviewers connected with *Zentralblatt* and *Mathematical Reviews* could be persuaded to organize and launch a range of conferences.

So as to enable users to search electronically for relevant computer mediated conferences, the description of a computer mediated conference, especially the subject area, should be defined in a manner that is suitable for searching, and the bibliographical information retrieval facilities of the system should be provided for searching computer mediated conferences.

New research areas are not initiated by proclamation from some official authority; they emerge. So do research networks. In order to accommodate for that, any EUROMATH subscriber should have access to create a computer mediated conference within the EUROMATH system.

Participation in normal research related conferences should be open to any EUROMATH subscriber. On the other hand, it should be possible to set up conferences with limited access, e.g. for editorial boards.

Some of the conferences will invariably generate a vast amount of communication (e.g. in some of the very large research areas and in topics of general interest). The system should provide facilities allowing the user to conduct automatic or interactive searches in the mass of messages in conferences. That is, the basic search facilities of bibliographical information retrieval systems must be provided for computer mediated conferencing as well. In order to support that, messages must have a 'header' providing information on the name of the sender, the date, the subject (keywords, possibly abstract), the size of the message, etc.¹¹

2.2.2. PREPRINT DISTRIBUTION

A vast volume of copies of unpublished mathematical papers is being distributed via personal channels and, to a lesser degree, via distribution centers. These so-called preprints can be divided into three types:

- (1) Papers that have been accepted for publication but at the given moment are still in the technical stages of the publication process.

¹¹ Cf. Lyne DaSylva: "First Field Study of Electronic Mail and Conferencing System", EUROMATH, Task 2.1.3.: Field Study of Interpersonal Communication, DDC, 1987.

- (2) Papers that have been submitted for publication but at the given moment are still under consideration by the editor and the referee.
- (3) Draft manuscripts of papers intended to be submitted for publication later.

The main reason for the high volume of circulation of preprints is the excessive lead time incurred by the production of journals. Typically, the lead time is from 1.5 to 2 years, in some cases even 3 years.

One of the biggest delays on the way to having a paper published is the time taken for refereeing the document. Once the document is submitted, the editor of the journal must find one or more competent referees. The editor usually knows of several potential referees in any given area of mathematics upon whom he or she may call. It may transpire that the people upon whom the editor calls are unavailable. This is frequently the case as referees also have teaching and research obligation. They may also be on the board of more than one journal and may loath to accept more than one paper at a time. That is, the search may have to be reiterated several times until, eventually, a referee is found.

Because referees provide an unpaid service to the community it is regarded as improper to exert pressure on referees to carry out the task quickly. Also, the task of critically reviewing a twenty or thirty page mathematical paper is far from trivial. For these reasons, the referee is given from two to three months to produce a report on a paper. Once the paper has been evaluated, a report is sent back to the editor who then advises the author as to whether the paper as it stands is acceptable for publication or needs further enhancements recommended by the referee. If the paper is accepted for publication it is still returned to the author for verification. If it is not passed then the author will be asked to make the necessary changes to it. If these are trivial amendments then the editor will usually be able to confirm the correctness of the new version of the paper without further reference to the referee. If the changes to be made by the author are more fundamental then the corrected version is passed back to the referee for further scrutiny. In some cases a further iteration is required before the paper is finally accepted. Not surprisingly then, it is not uncommon for the total process of refereeing a paper to take six months.

The following technical stages of the publication process - typesetting etc. - do not appear to contribute significantly to the total publishing delay time. Much of the delay at this stage is related to proof reading by the authors and this can at times be significant if there is more than one author involved.

Even after typesetting and proof reading substantial delays may arise, due to the policies of the publisher. Editors may, for example, stagger the publication of papers in order to maintain a certain uniformity of size and content over a number of issues. Accordingly, accepted and corrected papers may be placed in a queue while awaiting publication.

To circumvent these delays in the publication process, almost all mathematicians distribute to colleagues copies of papers that have been submitted for publication. The main channel of distribution is personal contacts, that is, the author sends his or her paper directly to relevant colleagues. This state of affairs is not satisfactory, however. It means that the objects of discourse of the mathematical community to a large extent are preprint papers and these are not normally publicly available.

Attempts to solve the problem have been made in the form of distribution centers like, for instance, the Royal Institute of Technology in Stockholm. The center acts as a repository and distribution channel for preprints. It issues listings of new preprints (author, title and abstract) to subscribers and accepts requests for copies. For economic reasons listings are only issued infrequently, e.g. three times a year.

The administrative costs of storing and retrieving copies and producing and distributing listings are high, and the lack of administrative resources causes periodical delays in the distribution of new lists.

To deal more effectively and efficiently with the issue of public access to the circulating mass of preprints, a full text database containing preprints of mathematical papers has been suggested. The findings of the requirements analysis indicate, however, that the feasibility of such a database is questionable.

(1) The number of preprint copies issued annually is of the same order of magnitude as the number of papers published annually, i.e. more than 35.000. Indeed, the number of preprint copies issued is greater than the number of papers published. Firstly, preprints are issued for papers that eventually do not get accepted for publication. Secondly, preprints may be issued in several versions. Since the data in a preprint database normally will be outdated in, on average, two years, the cost of entering the data can hardly be justified. Unless, of course, the data is entered directly by the authors or by their institutions, but do they have any incentive to do so, and will they be able to index the papers well enough? The experience of the preprint distribution centers are pessimistic. While authors are encouraged to include brief abstracts with their papers this is sometimes neglected. FIZ generally produce the abstract themselves.

(2) The quality of the papers in a preprint database could not be controlled by any editorial or refereeing institution. Accordingly, a preprint database may become an alternative unrestricted publication channel. The reaction of the mathematical community to such an alternative channel may be one of sound skepticism, or worse, the database may have detrimental effects on the quality of mathematical research.

(3) Even more importantly, an open and unrestricted distribution of preprints may simply be in discord with the competitive climate prevailing in most areas of mathematical research. As noted above, mathematical research is typically conducted within relatively narrow specialties within which multiple mathematicians are working on the same problem. This state of affairs generates a relatively competitive climate. The competitive climate can become fairly severe.¹² In the case of Poincaré's conjecture, for example, years of unrewarded dedication has given rise to intense competitive feeling among the involved scientists. When asked by Charles Fisher how he would feel if another person settled the conjecture, one of the informants made a very sad face and returned the question. Another replied:

A: I would feel horribly disappointed. To be honest, I take a dog-in-the-manger attitude to this problem.

Q: You mean you would just as soon not see it solved if someone else were to solve it?

A: That's right, that's right. I am sure that this is morally irresponsible, but that is the way I feel about it. I don't burn black candles or anything.¹³

This may be a somewhat extreme position caused by the intransigence of the problem and years of unrewarded dedication, but the competitive mood is pervasive. As observed by Richard Startup in a study in a British university:

¹² Cf. Randall Collins and Sal Restivo: "Robber barons and politicians in mathematics: a conflict model of science" (1983).

¹³ C. S. Fisher: "Some Social Characteristics of Mathematicians and Their Work" (1973), p. 1114 f.

“What distinguished the pure mathematicians from the academics in the three other selected departments [classics, civil engineering, and psychology] was that they were particularly conscious (and worried) that their research advances might be anticipated by others.”¹⁴

Actual anticipation is less prevalent in mathematics compared to physics and molecular biology, but the issue of anticipation is felt more acutely (cf. fig. 2). As suggested by Hagstrom, being anticipated is more likely to forestall publication by mathematicians than by experimental scientists. In experimental work like, for instance, determining physical constants, demonstrating new effects etc., replications tend to be desirable and papers reporting on replications are publishable. As pointed out by Hagstrom, this is less likely to hold true for mathematical theorems. Replications are not needed in mathematics.¹⁵

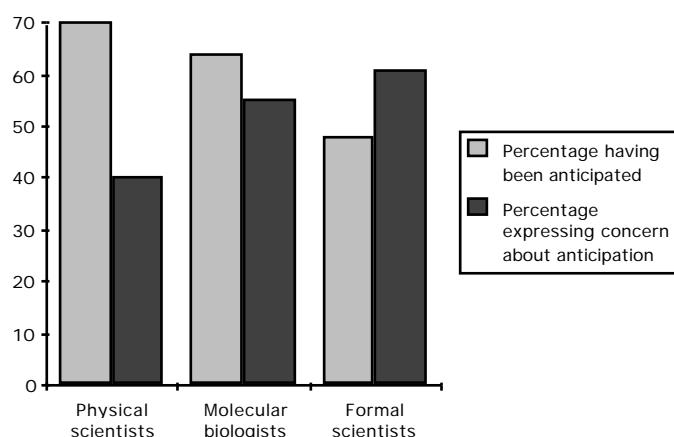


Fig. 2. Experience of anticipation versus concern about anticipation in three scientific fields: Physical scientists (theoretical and experimental physicists and physical chemists), molecular biologists, and formal scientists (mathematicians, statisticians, and logicians).¹⁶

Access to the preprint copy of an unpublished paper by one mathematician may be very beneficial to the research of another. The reader of the preprint might however have access to publication channels with a minimal lead time. Thus he or she may be able to publish his own, possibly more advanced, results before the paper that helped him along is published. Even though the reader of the preprint *may* pay due homage to its author by means of citations, the very fact that the gist of it has been made publicly available and has been surpassed may mean that the show is over. That is, by the time the paper is finally published the results may be outdated and appear well known. The originator of the solution or idea in question will not receive due credit.

In other words, preprints are normally very sensitive documents. Preprints are only mailed to colleagues considered to be trustworthy. Likewise, preprints that have been received on the basis of trust are not passed on to third persons. This is an unwritten codex that is generally adhered to in the mathematical community, as opposed to other scientific communities where copies of preprints are passed on from hand to hand without the authors knowledge or consent.

¹⁴ Richard Startup: *The University Teacher and his World. A Sociological and Educational Study* (1979), p. 70.

¹⁵ Hagstrom: *The Scientific Community* (1965), p. 76.

¹⁶ Hagstrom: *The Scientific Community* (1965), p. 75.

In some areas of mathematical research, however, competition is not as acute as in the more typical research areas. In those areas preprints seem to be distributed fairly unrestrictedly like in other disciplines.

In sum, because of the competitive climate prevailing in mathematics a separate full text preprint database may be expected to attract only a small fraction of the total mass of circulated preprints. Of course, if inclusion of a paper in the database was perceived as being identical to its publication, the fear of not receiving due credit for the fruit of years of toil would dwindle. But this means circumventing the refereeing process and implies its ultimate demise. This is, of course, absolutely not desirable.

Furthermore, since the number of preprints that will be made publicly available is small, the concerns stated above (§ 1) are of minor import. However, a database of the sort envisioned must be fairly comprehensive in order to be usable and acceptable. Thus a *separate and comprehensive* preprint database (whether full text or not) seems unfeasible.

A more realistic proposal, which is easier to implement, is to use existing facilities to more quickly provide information about those preprints that authors are willing to make publicly available before they are published. This simple solution to this non-critical problem could be implemented as follows:

It is proposed that whenever a paper has been accepted for publication, then either the journal or the author will immediately send the necessary information (Author, Title, Abstract, Journal issue, Classification, etc.) to FIZ, who will then register the paper in *Base Math*. (FIZ have already established this procedure for some journals.) This has the advantage that it makes the paper available to those interested before the journal is actually published, i.e. it cuts out delays due to typesetting and the queue of accepted articles awaiting publication. Mathematicians who learn of an article via *Base Math* can then order it via EUROMATH. It has to be decided whether FIZ will be willing to handle such orders, or whether it should be done by the author's institute or the author himself. In any case, this should be sorted out by the system - behind the user's back.

Each mathematical department (institute etc.) can start a preprint publication series properly registered with ISSN and ISBN (if they do not already have one; some departments already have such preprint series). According to their own discretion authors may print their submitted manuscripts in this series upon which they are then publicly available. Each issue of the series is to be submitted to FIZ and AMS for registration in bibliographic databases. In the case of a manuscript of a paper that has not yet been accepted for publication by the editorial board of a journal and which the author wants to make publicly available, it is proposed that the manuscript is published in the preprint series of the institute if the author wishes to make it publicly available. This way the bibliographic data on the preprint will enter the bibliographic databases.

In addition, a user can still send out a preprint, or a description of a preprint, to a list of mathematicians, via the electronic mail facilities of the EUROMATH Information Delivery and Retrieval System. Further, since EUROMATH will provide a list of all European mathematicians that includes their areas of interest, the author can exploit this information to also send it to people he or she does not know, but who might be interested. In this case, it is clearly the author who should receive orders for the paper.

Of course, the author could even go further by drawing attention to a paper in a EUROMATH electronic conference.

Finally, the author may choose to publish the full text of the paper, whether accepted for publication or not, as a contribution to a conference. The EUROMATH Information Delivery and Retrieval system should allow such a usage.

The proposal for handling preprints is summarized in fig. 3.

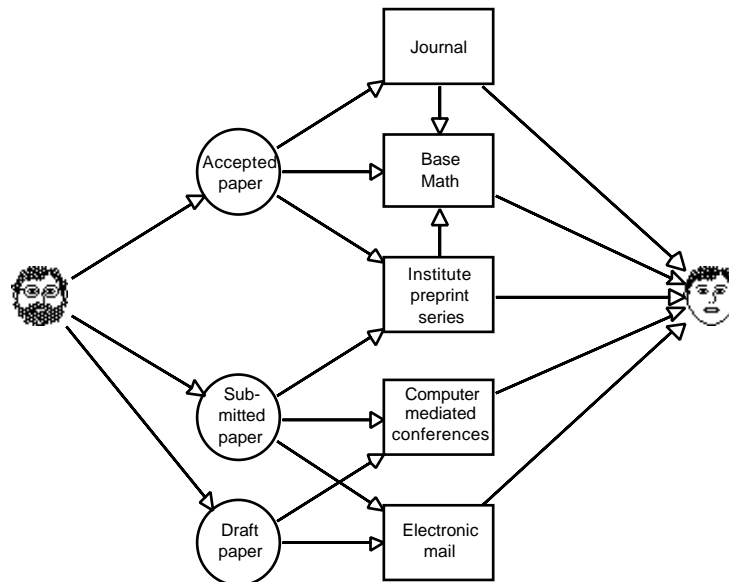


Fig. 3. A proposal for handling preprints in EUROMATH. The circles represent the three types of preprints and the boxes represent the various channels of disseminating bibliographic information on mathematical papers or the papers themselves.

2.2.3. ELECTRONIC PUBLISHING

It seems likely that gradually, over the next few years, mathematicians will more frequently submit papers to journals in electronic form, e.g. as a \TeX source file on a diskette. Eventually, it is foreseen that all journals will insist on this form of input, i.e. *electroniccollection*. As this practice becomes more established, it also improves the feasibility of *electronicdistribution*, i.e. of distributing journals on CD-ROMs, via electronic mail, or via on-line full-text databases. Again \TeX source files might be used, but \TeX output files and other representations are also possible. Electronic distribution is likely to become popular, as it will probably be cheaper than paper distribution, and once the necessary standards have become widespread, its disadvantages also become less significant.

The revolutionary feature of electronic publishing is that it does away with the *necessity* of employing a publisher in order to produce a journal, since it abolishes the printing process. Of course, a hard copy does have its advantages - both objective and subjective. So there may still be a role for a publisher (as a printer). There is also a question of who would perform the administrative tasks involved in producing a journal, which again may provide a role for publishers. However, the main point is that all that is really *necessary* to set up and run an "electronicjournal" is an *editorial board* and a *distribution channel*. EUROMATH can be such a distribution channel.

Of course, there is still some uncertainty as to if and when electronic publishing will come of age. So, the best strategy for EUROMATH is to make sure that the doc-

ument interchange standard and the distribution medium become established, and remain in the control of the mathematical community, so that it is ready to play a dominating role in electronic publishing.

2.2.4. BIBLIOGRAPHIC INFORMATION RETRIEVAL

As noted above, a mathematician who is well established in his field will tend to know most of the active researchers in this field, or at least he or she will know of their work. He or she will be engaged in regular correspondence with a large part, perhaps the majority, of the active scientists in the research area. He or she will send them preprints and reprints of his papers and he or she will receive preprints and reprints of their papers. In this way, mathematicians - especially those who have established themselves in the area - are continually being informed about new literature pertinent to their research. In addition, they will browse the last issue of a few chosen journals so as to monitor the area and watch out for interesting papers. As a rule, mathematicians express a high degree of confidence in the reliability of this very informal system of distribution. Considering this widespread confidence in the informal updating system and the high costs of accessing bibliographic databases, it is no wonder why mathematicians practically speaking do not use the existing electronic databases at all.

Consequently, the bibliographic information retrieval facilities to be provided by EUROMATH are up against strong competition in advance. The additional services provided by EUROMATH must be conspicuous and convincing if they are to get acceptance and the price of retrieving bibliographic information must be reduced substantially, if the cost-benefit trade off is to be changed decisively.

Some of the most important elements are discussed below:

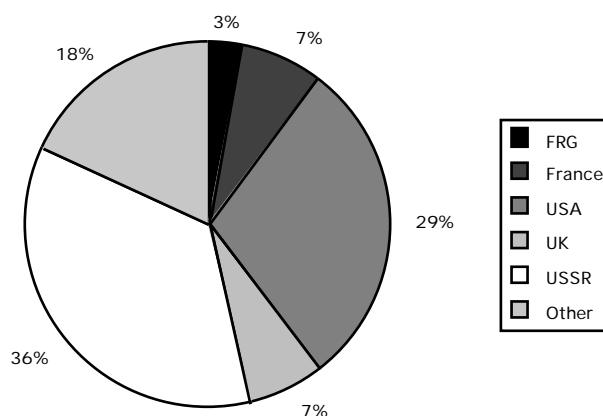


Fig. 4. Percent distribution of the world's mathematical articles by country of author, 1972.¹⁷

Mathematics is an intercontinental endeavor. (Cf. fig. 4). The bibliographic information retrieval facility of the EUROMATH system must provide access to bibliographic information pertaining to all major sources of mathematics, that is, all of the currently existing databases. It is particularly striking that a very large part of

¹⁷ Schøtt: *Fundamental Research in Small Countries. Mathematics in Denmark, 1928-1977* (1979), p. 98.

mathematic papers are produced in the USSR. Access to comprehensive information on the literature published outside of Western Europe and North America would be a strong asset.

As noted above, mathematics, as opposed to, say, physics, is a fairly 'flat' science. Accordingly, terminology is diffuse across specialities. What is essentially the same phenomenon may have very different names in the different research areas as well as in the different countries. Therefore the EUROMATH system should include a thesaurus to support the user in obtaining an acceptable degree of retrieval effectiveness.

The discourse of the mathematical community, as any scientific discourse, revolves around papers and other types of publications. In fact, the products of mathematical research are embodied in papers and nothing but papers as opposed to, for instance, specimens, isolated substances, recordings of experimental data, etc.

Accordingly, a facility should be created within the EUROMATH environment that will allow users to trace subsequent references to a particular paper in the databases of the EUROMATH system by means of queries like, for example, 'Which conference contributions have citations of this paper?'. This facility can be created as a cross-reference facility within the EUROMATH Information Delivery and Retrieval system by integrating the bibliographical retrieval facility and the interpersonal communication facilities. References to a paper could be based upon the ID numbers of the records within the FIZ and AMS bibliographical databases. When a user wants to comment on or cite a paper, he/she can invoke a special command in the front-end of the EUROMATH Information Delivery and Retrieval system which can then acquire the ID number from the bibliographical databases (or locally stored subsets of those databases, downloaded in previous sessions) automatically, or semi-automatically, and add the number to the text as an embedded command.

Of course, the cross-reference facility should also enable a user to acquire full bibliographical data of a particular reference in a contribution to a conference and, if needed, to procure a hard copy of the reference *via* the database host.

Likewise, it would be very beneficial if it was possible to search on citations by means of queries like, for example, 'Which papers have references to this paper?'. However, this facility would demand major additional resources for the task of entering new data in the bibliographical databases - unless papers are generally available in electronic form in a high-level, structured representation - and for updating existing data.

Finally, a service automatically providing subscribers with a monthly update on new literature within a user-defined research area would seemingly be received positively by mathematicians. It would fill a need. The service should be offered to all irrespective of whether they have access to computer retrieval facilities or not; the update may be distributed by ordinary mail or transferred electronically to the user by means of the EUROMATH Information Delivery and Retrieval System. This service is already provided by most on-line database hosts. However, the service should be marketed very actively and priced very aggressively so as to attract, hopefully, the majority of mathematicians. This arrangement would presumably increase the degree of utilization substantially with no additional cost and it would, perhaps most importantly, provide precise and fresh data on the research profiles of the subscribing mathematicians to be entered into the EUROMATH directory.

2.3. The EUROMATH Directory

In order to enhance cooperation within the European mathematical community, a specialized EUROMATH Directory is needed. The system should give access to descriptions of the following entities. (A more complete description of the data to be stored is given in the Requirements Specification.)

- Individual mathematicians.
- Mathematical institutes.
- Mathematical associations.
- Mathematical meetings, i.e. conferences, workshops, etc.
- Sources of funding for mathematical research.
- Mathematical journals.
- EUROMATH computer mediated conferences.

The system must allow the user to search for any combination of data. The query language (as seen by the user) must be identical to that used for biographical information retrieval. It should allow the user to combine criteria on different attributes of different entities, e.g. to search for the names of those mathematicians in a given country that are active in a given research area.

In view of the fact that the informal research networks are of a genuinely inter-continental scope, the database of the EUROMATH Directory should cover the mathematical community, not only in Europe, but on a world wide scale.

2.4. Personal Databases

The EUROMATH Information Delivery and Retrieval System must allow the user to file and retrieve data for personal use, e.g. files of mail received and sent, mailing lists, bibliographies and exam results.

Some of the information must be accessible by other EUROMATH users; hence it must be filed within the direct jurisdiction of the EUROMATH Information Delivery and Retrieval System. For instance, by storing mailing lists within the jurisdiction of the EUROMATH Information Delivery and Retrieval System it is possible to have mailing lists updated automatically, e.g. upon change of address. By storing search profiles within the jurisdiction of the EUROMATH Information Delivery and Retrieval System they can also be used for defining the research areas of individual mathematicians.

Furthermore, the system must allow the user to file and retrieve information (e.g., correspondence, bibliographic search results) locally, that is, beyond the jurisdiction of the EUROMATH Information Delivery and Retrieval System. The interface to the filing and retrieval procedures of this facility must be totally consistent with the interface to the filing and retrieval procedures of the EUROMATH Information Delivery and Retrieval System.

2.5. The EUROMATH Information Delivery and Retrieval System

To summarize, in order to support the discourses of the mathematical community it is proposed to provide the following facilities in EUROMATH:

- Person-to-person(s) letters (electronic mail).
- Computer mediated conferences:
 - Open (i.e. accessible to any user) or closed (i.e. restricted to a specific group of users).
 - Moderated (all contributions are first routed to a particular user, called a moderator, who then makes sure that only relevant contributions are passed on to the audience of the conference) or non-moderated.
- Electronic journals, which are akin to moderated conferences except that the moderator is replaced by an editorial board who are responsible for the refereeing of the received papers, just as in traditional paper journals.
- The EUROMATH Directory.
- Bibliographical search of review databases.
- Cross-referencing of references in contributions to computer mediated conferences and bibliographical databases.

As a whole, these facilities constitute the EUROMATH Information Delivery and Retrieval System.

The proposed architecture of the EUROMATH Information Delivery and Retrieval System is illustrated in fig. 5. A front-end application program running on the user's workstation and a back-end application program running on the EUROMATH backbone computers provides uniform access to the various database facilities of the system.

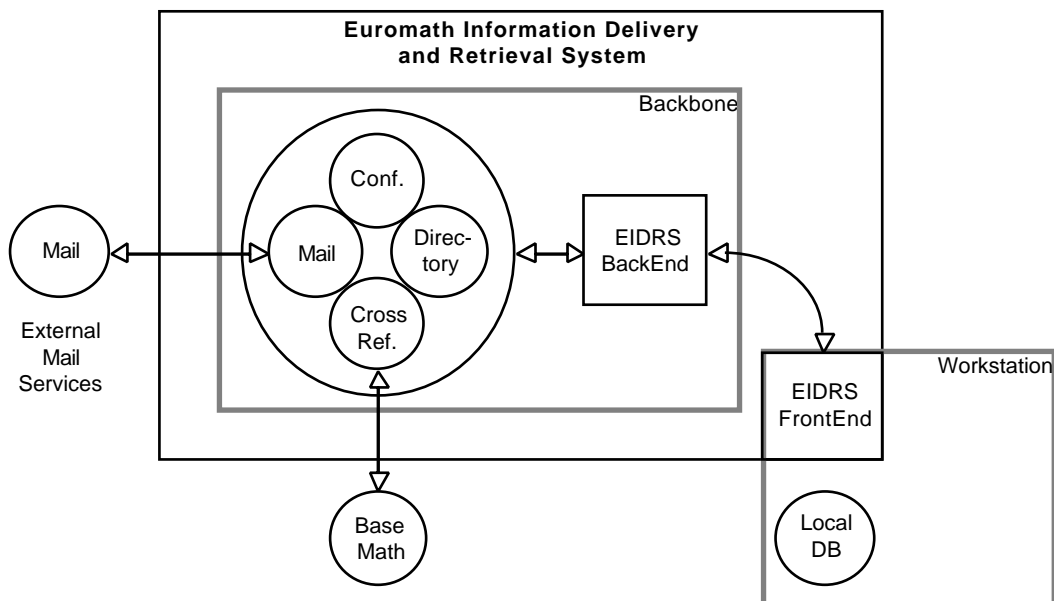


Fig. 5. The architecture of the EUROMATH Information and Retrieval System.

By providing uniform access to the various database facilities and by establishing cross-references, the EUROMATH Information Delivery and Retrieval System will be able allow the user to navigate relatively freely in a 'information space' encompassing the information world of mathematicians. (Cf. fig. 6).

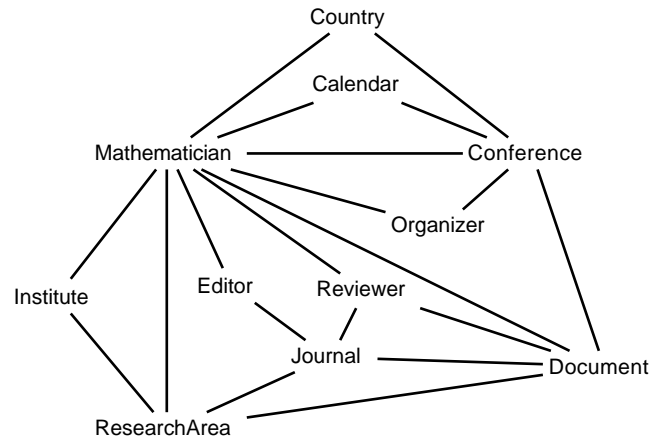


Fig. 6. The information world of mathematicians (simplified data model).¹⁸

¹⁸ For a more complete model, cf. G. McAlpine and K. Schmidt: "Requirements Specification", EUROMATH/DDC09/1988-10-13, chapter 4.

3. Document Preparation

The most conspicuous characteristic of mathematical papers is, of course, the abundant use of symbolic notation.

“The principal functions of a symbol in mathematics are to designate with precision and clarity and to abbreviate. [...] In point of fact, without the process of abbreviation, mathematical discourse is hardly possible.”¹⁹

By extinguishing the eloquent redundancy of ‘natural language’, mathematical discourse is utterly dependent on succinct graphical representation. Therefore, the typographical layout of mathematical expressions is of critical importance to the effectiveness of the discourse. It is well known, however, that the typographical processing of mathematical expressions is very resource demanding. This slows down the research process and it is a boring task. What is worse, it disrupts the iterative and continuous mode of working that is the natural in problem solving work. Computer-supported document processing should, and can, alleviate this problem.

3.1. Interface Issues

The mathematician does not stop thinking on mathematics when he or she is writing a paper. The process of constructing a correct, concise and lucid exposition is part of the research process. A document processing system for mathematicians must be designed so that it does not distract attention from the content of the task. It should allow the mathematician to keep his mind on the mathematics.

3.1.1. DIRECT MANIPULATION

The command language of the editor must correspond directly to the user’s mental representation of mathematical expressions. According to numerous statements from mathematicians, \TeX corresponds very well to mathematicians’ mental representation of mathematical expressions. In terms of a user’s mental model, then, \TeX is a suitable candidate for the text exchange format.

3.1.2. WYSIWYG

The process of formulating thought and constructing an exposition is an iterative process. The interim formulations will often give rise to reconsiderations. The author may want a better expression or may suddenly realize that the structure of the argumentation is suspect. The document processing system should therefore continually display to the user the current appearance of the text when it is printed (‘What You See Is What You Get’ or WYSIWYG). That is, the system should allow the user to see the typeset text as well as the mark-up (or other form of) source code - preferably simultaneously in two windows.

¹⁹ Davis and Hersh: *The Mathematical Experience* (1981), pp. 123 f.

3.1.3. INTERACTION STYLE

Likewise, to allow the user to keep his mind on the mathematics the system should allow the user to enter and format text and make changes in both modes. That is, the user should be able to work in the WYSIWYG window if he or she so desires.

At least for conventional document processing tasks, the system must allow the user to format text in the graphic style of interaction of modern document processors, i.e. by providing menu choices of fonts, styles (italic, bold etc.), font sizes, indentation etc.

3.2. Alphabets and Fonts

In order to be accepted by users and in order to facilitate exchange of documents between mathematicians in different countries the document preparation system must support all European alphabets. That is, (1) it must support the letters and diacritical marks of the various Latin alphabets, e.g.: ç, é, à, â, ñ, ß, , ä, ö, æ, ø, å, c]r] ɾ, etc., and (2) it must support the Cyrillic, and Modern Greek as well as the Latin alphabets.

Furthermore, the system must support the complete repertoire of mathematical symbols, the Classical Greek alphabet included.

The repertoire of mathematical symbols is open ended. It is perpetually evolving:

“The mathematical font of special symbols currently in common use comprises several hundred symbols with new symbols created every year.”²⁰

Consequently, the system must allow for the user to design mathematical symbols to be included in the fonts of the system.

3.3. Document Structure

The system must provide facilities for automatic formatting of the specific structural elements of mathematical documents:

- definitions;
- propositions;
- formulae;
- theorems, etc.

The system must know the pertinent structural elements and the particular formatting of each of these elements. The system should provide a standard default option (e.g., L^AT_EX), but the user should be able to override the formatting definitions as well as to add other elements. The facility corresponds to the style sheet facilities of modern document processing products.

Furthermore, the system must provide facilities for automatic generation and handling of the following general structural elements of documents:

- pagination;
- footnotes;
- endnotes;

²⁰ Davis and Hersh: *The Mathematical Experience* (1981), pp. 122 f.

- references;
- table of contents;
- table of figures;
- tables of definitions, propositions, theorems;
- indexes;
- headers and footers, etc.

The facility must encompass automatic (albeit user definable) formatting of the various automatically generated structural elements.

The system should enable the user to change the layout of the document as a whole as well as the format of any of the structural elements of the document. That is, it should provide a so-called style sheet facility.

3.4. Outlining

Naturally, mathematicians do not postpone writing until they sit down to produce a paper or a book. Writing is an integral part of the continuous and iterative research process. During that process the mathematician will make a list or a stack of notes: conjectures, definitions, equations, diagrams, examples and counter-examples, lemmas, theorems, and sketches of proofs.

In the words of Richard Palais:

“A paper evolves and is written from its inception to its completion in many small steps, better reflecting the way the writer actually carries out research, rather than in a sequence of discrete, separate, and major rewrites that in fact interrupt the process of research.”²¹

Accordingly, the document processing system should provide a so-called ‘outlining’ facility, i.e. a facility for interactive generation of hierarchical text structures like, for instance, an outline, a synopsis etc. of a document.

By allowing the user to enter notes more or less randomly and very easily make radical changes to the structure of the document, an outlining facility supports the continuous and iterative mode of working very efficiently and effectively.

The outlining facility should be fully integrated with the editor. The outline should be merely another way of viewing the text file. It must be possible for the user to define the various structural elements (definitions, theorems etc.).

3.5. Illustrations

In some mathematical research areas, graphics plays a key role. In those areas, the inclusion of graphic illustrations in papers is a *sine qua non*. Accordingly, the document preparation system should allow the user to include computer-generated graphics in papers in electronically stored format in a convenient way. For example, to facilitate the transfer of data from application programs generating graphics to the document processing system the operating system of the user’s workstation should provide a simple ‘cut and paste’ function (like that of Apple’s Macintosh).

Formatting of graphic objects within the document processing program should be at least as direct and interactive as in modern desk top publishing programs. For

²¹ R. S. Palais: “Mathematical Text Processing” (1986), p. 746.

example, the system should be able to generate the corresponding T_EX code automatically.

It should be possible for the user to resize (crop, reduce, enlarge) the graphic objects.

In accordance with the requirement for WYSIWYG, the system must display the appearance of the page with graphics in place.

3.6. Other Facilities

For the document preparation system to be acceptable to users having experienced the amenities of modern document processors, other features are required as well:

(1) The document preparation system should include, or allow for the inclusion of, spelling checkers in multiple languages. The spelling checkers should be able to run in batch mode as well as interactively.

Preferably, the spelling checker should include, or allow for the inclusion of, an interactive thesaurus in each of the supported languages.

(2) The document preparation system should also include, or allow for the inclusion of, a hyphenation function operating in multiple languages. The hyphenation function should be able to work in batch mode as well as interactively.

For a fairly complete analysis of users' requirements in relation to document preparation, cf. the report from the Boston Computer Society.²²

²² Goldstein, Loomis, and Tetewsky: "Technical Wordprocessors for the IBM PC and Compatibles, Report by the Boston Computer Society. Part I TWP Capabilities and People Needs" (1987).

4. Research Support Tools

4.1. Symbolic Manipulation

Mathematicians are, of course, first and foremost occupied with the production of new mathematics, and only secondarily with communicating their results to their colleagues in the form of memos, papers, etc. A lot of attention is invested in preparation of typographically appealing and lucidly structured papers, not because that is the essential thing but because the papers represent the proud results of months or perhaps years of arduous work.

In other words, the need for computer support cannot be limited to the preparation and delivery of documents. The resources of mathematicians in the research process itself need enhancement. A steady proliferation of symbolic manipulation systems is most likely.

Accordingly, it is sound to state that the mathematical community have a great latent need for powerful workstations.

4.2. Graphics

In some mathematical research areas the production of graphics is crucial, not only for didactic reasons but also for the research process itself. The EUROMATH system must cater for the need to exchange graphics, both in the role of illustrations in papers and in the role of autonomous research documentation (e.g., in electronic mail).

The kind of graphics produced in mathematical research is almost always based on curves as opposed to grey tone graphics, e.g. graphs, net structures, etc. Accordingly, the Postscript page description language developed by Adobe, is a very adequate language for device independent definition of graphics and, hence, for exchanging graphics within the EUROMATH system.

5. Integration

Perhaps the most important idea behind EUROMATH is that a set of facilities based on information technology already exist for mathematicians, but that they are not used enough because they are too expensive and too hard to use. Further, one of the main reasons for difficulty of usage has been assumed to be lack of integration.

By integration is meant that the facilities that can be accessed fit well together, e.g. transfer of data and acquired skills should be feasible and easy.

5.1. Different Types of Integration

To discuss integration, one must distinguish between (at least) three types of integration:

- Inter-workstation integration: data can be transferred from one workstation to another, irrespective of the platforms involved.
- User interface integration: the system allows the user to perform similar tasks in a similar way in different application programs.
- Application integration: the semantics of data transferred from one application program to another is retained for further processing. The user perceives the given collection of application programs as a unitary application environment.

5.1.1. INTER-WORKSTATION INTEGRATION

Inter-workstation integration can be achieved by means of fairly low-level standards, because it is basically a question of transferring data between essentially identical applications (e.g., document processor to document processor) operating on different platforms (e.g., DOS/MS-Windows, Unix/X-Windows, Macintosh, etc.).

On the basis of the preliminary findings of the Requirements Analysis, the following EUROMATH inter-workstation data exchange standards seem feasible and appropriate.

- Text interchange format: Ordinary ASCII (i.e., text files stripped of commands) and \TeX .
- Document interchange format: \LaTeX (incl. \TeX).
- Graphics interchange format: Postscript.
- Embedded graphics in text: Encapsulated Postscript Format.

It is proposed that a decision be taken to the effect that all documents to be filed in the EUROMATH system (e.g. contributions to computer mediated conferences) must adhere to these standards.

5.1.2. APPLICATION INTEGRATION

Of course, the ideal design of the EUROMATH system would be a comprehensive integration of application programs that would allow the user to command the entire range of functionality of the system within a unitary application environment; that

is, an environment where the structure and segmentation of the software system is invisible to the user.

It is always possible to transfer data from one program to another in some way. As a last resort, one can always write a program to do so. So, the relevant question is: what degree of data integration can feasibly be achieved across applications?

A satisfactory degree of data integration is obtained if the user, by using a single command, can copy data from any of the integrated application programs to a temporary repository (e.g. a file or a clipboard), and then, from any of the other integrated application programs, can directly import this data from the temporary repository, using a single command. To fulfill this requirement, some standard data representation for data exchange must be defined and adhered to by the application programs.

Application integration requires a common representation of the semantics of the data. For example, a user switching from a spreadsheet facility to a document processing facility should not only be able to transfer the data (the values of the cells); the relationships between the cells, as defined in the spreadsheet, should be retained as well and should be accessible from the document processing facility.

5.1.3. USER INTERFACE INTEGRATION

It is obviously a crucial requirement that we provide a satisfactory degree of user interface integration across application programs. A convention must be designed and enforced. This requirement poses no major problems so long as the various application programs are designed by the same organization. Thus, it is perfectly feasible to enforce such a convention on all application programs developed within the EUROMATH project.

5.2. Integration vs. Flexibility

We must assume, however, that EUROMATH cannot develop and maintain a complete system that will cover all the various needs of mathematicians for computer support. Therefore, to be flexible the EUROMATH System should be extendible as well as integrated. By extendible is meant that it should be a trivial task for the user to integrate the EUROMATH system with all the other programs that the user wishes to use on the same platform.

In order to be extendible, but remain integrated, the EUROMATH System must be open, that is, it must have a well-defined set of interfaces that provide a means of integrating application programs. These interfaces can then define the required standard data representations. If a system were to provide complete user interface and data integration, then this would require a set of standards for both user and data interfaces. The system architecture illustrated in fig. 7 is highly suitable for such a system. Such a system is *open* in the sense that the interfaces between application programs and the other components of the system are well defined and published, so that if other application programs are developed to adhere to these interfaces, then they will also be integrated.

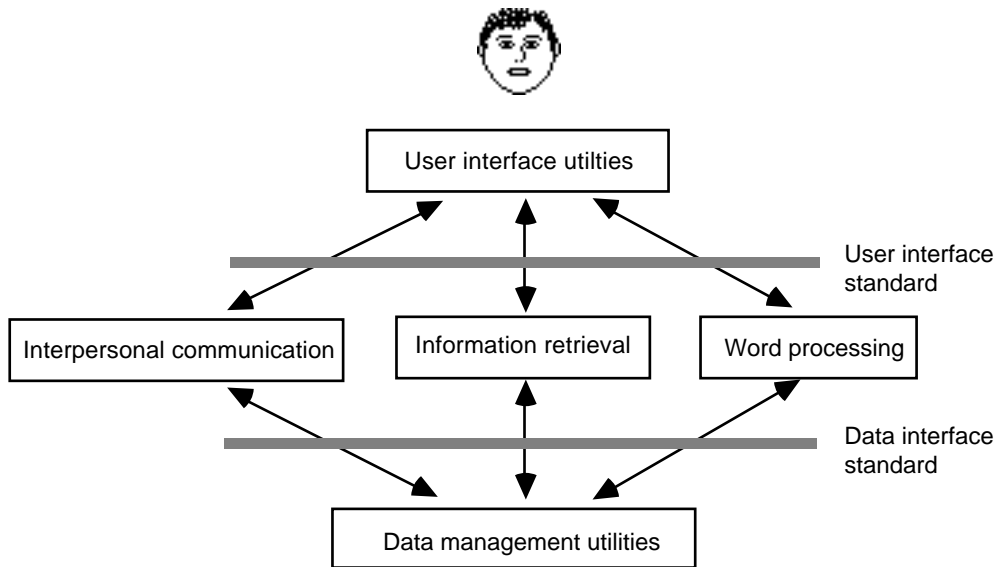


Fig. 7. A system architecture that ensures a high degree of integration because all application programs are required to use common user interface and data management systems.

The above architecture ensures a high degree of integration because all application programs are required to use common user interface and data management systems. However, a satisfactory degree of data integration can also be achieved with less sophisticated architectures by defining a user interface convention to which each application program should conform, a set of data exchange standards, and a 'hook' in each application program for transferring data using some data exchange mechanism. A simple architecture ensuring a satisfactory degree of data integration is illustrated in fig. 8.

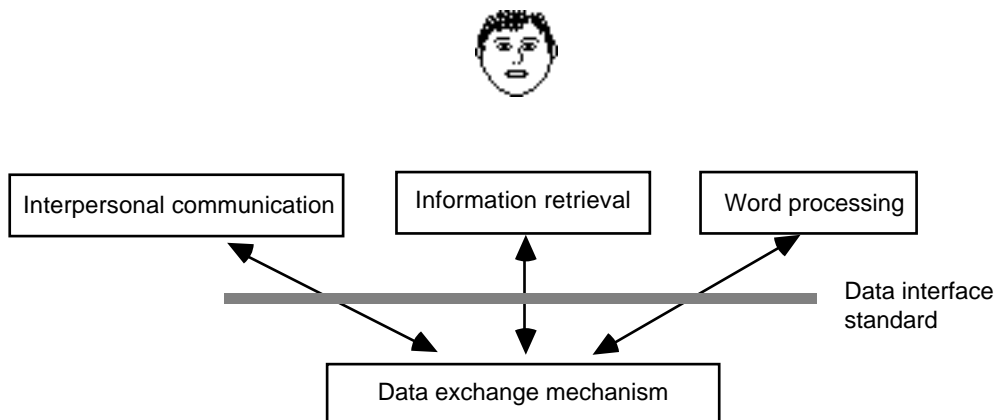


Fig. 8. A simple architecture in which a satisfactory degree of data integration is still possible. All that is required is a set of data exchange standards, and a hook in each application program for transferring data using some data exchange mechanism.

By a hook is meant a facility of a program whereby data can be output to another program and data generated by another program can be input. Most application programs have the capability to transfer their data, in ASCII form, to and from operating system files. This is generally the approach taken under MS-DOS, although it is not enforced in any way. Some platforms, like the Macintosh, make this easier

by providing a Clipboard via which data can be transferred using standard Copy and Paste operations.

Application integration would require another, more advanced implementation of 'hooks' by means of inter-process procedure calls. In a completely integrated, object-oriented system, the user does not need to think of transferring data at all. Whenever a particular object is selected, all the operations applicable to it, regardless of the situation, are made immediately available. For example, if a formula was selected, it would be possible to change its layout, simplify it or change the formula itself.

So the minimal condition for providing an extendible, integrated work environment are the presence of the necessary hooks and data exchange standards (as well as user interface conventions).

In practice, we are generally dependent on the existence and acceptance of appropriate standards and conventions. Unless we expect that other manufacturers will conform to our standards, the realistic degree of integration of external application programs within EUROMATH will be entirely determined by whether appropriate standards exist and are widely adopted outside of EUROMATH.

For example, suppose the user wants to integrate a computer algebra or spreadsheet program with the EUROMATH System. Then, unless the manufacturer of these systems adheres to the user and data interfaces we have defined in EUROMATH, integration is impossible. That is, one cannot transfer data from the computer algebra system to the EUROMATH System unless the EUROMATH System is specifically programmed to translate the external system's data format into the EUROMATH representation. Similarly, one cannot transfer data from the EUROMATH System to the computer algebra system unless either the computer algebra system provides a general hook for this and the EUROMATH System translates data into the external system's representation, or the external system is modified to accept data in the EUROMATH format. Thus, assuming our influence on manufacturers is limited, the integration of a computer algebra programs within EUROMATH would require that, for each program to be integrated, specific knowledge of these systems was programmed into EUROMATH. This does not fulfill the requirement for a satisfactory degree of extensibility, however.

Since we have to accept that the degree of integration that can be achieved depends on what standards have been adopted by the programs we want to integrate, it is essential to consider which data exchange standards have been generally adopted.

The suitability of a standard data representation will depend upon the semantics of the data that are to be transferred. In EUROMATH, it should, in principle, be possible to exchange all the data from the mathematician's information world that are stored electronically, cf. the data model specified in "A Model of the Mathematician's Information World" (Appendix). At the moment, the kinds of data that we can foresee the users wishing to transfer in EUROMATH include:

- (1) Unstructured, unformatted text, for which a feasible standard is ASCII.
- (2) Unstructured, formatted text, for which a feasible standard is $\text{T}_{\text{E}}\text{X}$.
- (3) Structured, formatted text, for which a feasible standard might be $\text{T}_{\text{E}}\text{X}$ or $\text{L}_{\text{A}}\text{T}_{\text{E}}\text{X}$.
- (4) Graphics, for which the most feasible standard might be Postscript.
- (5) Electronic mail addresses, for which a feasible standard is *user@site.domain*.

- (6) Mathematical formulae, for which there is currently no established standard.
- (7) References to documents (and perhaps other objects) accessible via the EUROMATH System, including published papers, preprints, mail/conferencing messages, for which no suitable standard exist.

Of these, all but the first two and Postscript must be considered so-called higher level representations.

Since the standards suggested for the representation of the first five kinds of information are already well established, their recommendation by the EMT should go down well with most mathematicians, and speed up their adoption by the mathematical community. This in turn will make it more worthwhile for manufacturers to support them, and if they do so, this will in turn increase the advantages of using them.

The fourth kind of information, graphics, is perhaps more complicated than the others. There exists several graphics standards, some of them at different levels. Postscript is at a fairly low level. An important issue in the future will be integration with computer geometry systems. However, since such systems are very rare at the moment, there will go some time before the problem of integration and a solution to it emerge.

Concerning representations for the sixth kind of information, i.e. mathematical formulae, different computer algebra systems use different representations, and the same thing applies for numerical computation systems. If no acceptable standard does exist, or is on the way, then it is not feasible to construct the EUROMATH System to provide a satisfactory degree of integration with computer algebra and numerical computation systems. The closest that we could get is to provide all the programs within the EUROMATH System with the capability to transfer formulae in a variety of representations, corresponding to the most commonly used computer algebra and numerical computation systems. It would then be up to the user to explicitly choose which representation he or she required. Furthermore, it is very doubtful if it is even possible to define a universal standard for the representation of all mathematical formulae. At best, one could define a standard for the most established mathematical notations and allow users to extend this for their own use. Thus, documents that included symbols from a user's own extension would not, in general be exchangeable.

Regarding the seventh kind of information, i.e. references, the situation is similar. However, this is not so serious a problem, since it is mainly within the EUROMATH System that this requirement is important.

Since we will not be able to provide all the facilities that the EUROMATH users require on their workstation as part of the EUROMATH System, the system must be designed accordingly. In particular, there seems little point in providing a special EUROMATH command interpreter if the user is going to have to learn the command interpreter of his workstations operating system anyway, in order to invoke application programs that are not part of EUROMATH.

Concerning extendible integration, we are severely bound by existing standards and platforms, which are rather inadequate for the comprehensive integration that we would like. Where they exist we should adopt existing, established standards, upon which we have no influence. Where there are no standards, we must either accept a lack of integration or try our own ad hoc solutions, which require a lot of maintenance. We can still design and implement an integrated, open EUROMATH System, but, in general, we do not have enough influence to be able to expect that

manufacturers will develop their programs to be compatible with EUROMATH. The exceptions to this generalization may be programs aimed specifically at the mathematical community, e.g. computer algebra systems. That is, where standardization has not already taken place, we could try and promote standards for the representation of data specific to mathematical applications, such as formulae.

Integration is a high priority requirement, but even more so is flexibility. The computer facilities to support the activities mathematical community must be adapted to highly diversified needs and preferences and must be able to integrated with an ever changing ensemble of application programs developed by other parties.

The precise balancing of the competing requirements of integration and flexibility calls for a more detailed discussion of the requirements of the various tasks involved.

5.2.1. DOCUMENT PROCESSING

The variation and differentiation of the needs and preferences of individual mathematicians in relation to document processing is immense. As observed by Richard Palais on the basis of countless reactions from readers of his column in *Notices of the AMS*:

“I noticed little uniformity in the point of the research and write-up cycle at which people claim to bring word processors into play. The methods mathematicians employ to help them create and communicate concepts and ideas are apparently too personal and variable to be easily codified.”²³

To cater for most of the wide variety of pertinent needs and preferences of mathematicians by means of one and only one document processing system is extremely ambitious.

More than that, it is not likely to be really desirable. The facility would be very complex and unhandy; unless, of course, the system was designed in a highly modular fashion with a generic document processing system and a large selection of specific functions from which each user can pick precisely those functions that he or she needs and prefers and then simply install them in the generic document processing system. This kind of design is not yet state of the art, however. And so far, adaptation to individual needs and preferences must primarily take the form of a pluralistic state of affairs.

Finally, the market for document processing systems is currently very dynamic and innovative. Good new systems are being launched at a very high frequency. The brilliant new systems of last year are already somewhat dated compared to the recent newcomers. To enter that market requires commercial courage and capital, not to mention what it takes to maintain a monopoly in a market like that.

If the EUROMATH system were to be based upon a particular document processing system, that is, if the other elements of the systems were to be designed to depend on a particular document processing system, EUROMATH as a whole would depend on the survival of that particular document processing system. In other words, an architecture based on a particular document processing system is not advisable.

²³ R. S. Palais: “Mathematical Text Processing” (1986), p. 746.

The architecture must allow the user to use any document processing system, provided the system is capable of producing the standard text and document formats (presumably $\text{T}_{\text{E}}\text{X}$).

5.2.2. SYMBOLIC MANIPULATION AND DOCUMENT PROCESSING

The foreseeable widespread use of symbolic manipulation systems in many mathematical areas raises a need for integration between symbolic manipulation systems and document processing systems.

Firstly, it is very desirable for a mathematician using a symbolic manipulation system to be able to automatically translate a formula written in the language of the symbolic manipulation system to the language of the document processing system (e.g. $\text{T}_{\text{E}}\text{X}$). Such a facility would save a lot of tedious typing and, more importantly, a lot of typing errors in complex formulas. Technically, this requirements poses no problems. However, as observed above, no standard representation for symbolic manipulation is emerging, and its emergence is unlikely due to the diversity of mathematics as a scientific discipline. To the contrary, most likely a wide variety of symbolic manipulation systems addressing different subdomains of mathematics will emerge. Thus, a series of ad hoc conversion programs are needed to cater for the transfer of formulae from the diverse and changing supply of symbolic manipulation programs.

Secondly, however, it may occasionally also be desirable for a mathematician to be able to make the opposite translation, for instance, to have a formula expressed in $\text{T}_{\text{E}}\text{X}$ translated into the language of the symbolic manipulation system. Thus it would be possible, for example, for a mathematician in the process of writing some research notes or a draft of a paper to investigate very easily whether an expression can be reduced. However, the nature of mathematical symbolic notation is an obstacle to fulfill this need.

“Whether or not symbols can, in fact, achieve absolute sharpness, fidelity, and unambiguousness is a question which is by no means easy to answer. At the end of the nineteenth century various committees were set up for the purpose of symbol standardization. They achieved only limited success. It would seem that mathematical symbols share with natural languages an organic growth and change that cannot be controlled by the ukase of a committee.”²⁴

The semantics of each symbol or symbol string cannot always be defined unambiguously. Indeed, the semantics of a given symbol or symbol string is highly context dependent. Even though mathematics is a formal science *parexcellence*, the language of mathematics is not formalized.²⁵

Now, it is perfectly possible to design a general language for describing mathematical formulae, a so called abstract syntax.²⁶ An abstract syntax of mathematical formulae would make it fairly easy to develop programs allowing conversions between different mathematical text representation standards. However, the set of primitives of such a language would be open ended. The language would have to

²⁴ Davis and Hersh: *The Mathematical Experience* (1981), p. 124.

²⁵ The fate of the logicist and formalist programmes in 20th Century mathematics (Russell, Bourbaki, Hilbert, etc.) has shown that mathematics cannot be formalized. (Cf. Davis and Hersh: *The Mathematical Experience* (1981), pp. 318 ff.). In 1931 such programmes were defeated in principle by the demonstration, by Gödel, that any consistent formal system strong enough to contain elementary arithmetic would be unable to prove its own consistency. (cf. Nagel and Newman: *Gödel's Proof* (1959).

²⁶ Cf., for instance, Arnon, Beach, McIsaac, and Waldspurger: “CaminoReal: An interactive mathematical notebook” (1988).

be expanded and modified constantly, in line with the progress of mathematical knowledge. More importantly, however, such a language would not retain and convey the *mathematical semantics* of the data being transferred. The semantics of any symbol and string of symbols would have to be defined by the user. The only way to retain the mathematical semantics of a given expression is to retain the source code of the symbolic manipulation program and the pointers from the elements of the typographical expression to the corresponding elements of the source code. Such a system is technically feasible (witness, for instance, Camino-Real), but it is essentially a closed system; that is, the mandatory formalization is hinged upon a local and temporary closure. Any given symbolic manipulation system is only applicable within a limited subdomain of mathematics. A universally applicable symbolic manipulation system is as unlikely as the transformation of mathematics into a unified, infallible and demonstrably consistent axiomatic system attempted by logicism and formalism in the first decades of this century.²⁷

Accordingly, a general facility for automatic translation of a mathematical expression in T_EX to a symbolic manipulation language is not feasible. The only feasible solution seems to be to provide access for users to define the semantics of a given symbol or symbol sequence. A system providing these facilities is hardly yet state of the art, however. And what is more, the representation would be parochial, that is, it would only be valid locally. The semantic representation would not be generally exchangeable within the mathematical community. That is, in case of integration of symbolic manipulation versus document processing systems application integration could only be achieved at the cost of flexibility and inter-workstation integration.

5.2.3. INFORMATION DELIVERY AND RETRIEVAL

The interfaces to (a) the bibliographic information retrieval facility, (b) the interpersonal communication facility (electronic mail and conferencing), and (c) the databases of the EUROMATH Directory, must be unified.

The user must be able to express the filter to be applied to computer-mediated conferences in the same language as the bibliographic search profile and the queries to the EUROMATH Directory.

5.3. A Feasible EUROMATH Architecture

In stead of a conclusion fig. 9 illustrates the possible EUROMATH architecture that emerged during the discussion above.

The EUROMATH Information Delivery and Retrieval System is a virtual center, probably distributed, providing uniform modes of access to foreign databases, the EUROMATH databases, and to EUROMATH subscribers.

The document processing system residing on the workstation is not an integrated part of the EUROMATH system; it may be any commercial document processing system that is able to input and output the standard exchange formats (T_EX).

²⁷ In fact, the philosophical platform of projects like CaminoReal is that of the defeated logicist and formalist programmes. Having been defeated in mathematics, the exiled logicist and formalist programmes have taken refuge in computer science.

The Information Delivery And Retrieval Front-end is part of the EUROMATH system. It is a front-end application program taking care of all communication with and through the EUROMATH Information Delivery and Retrieval System.

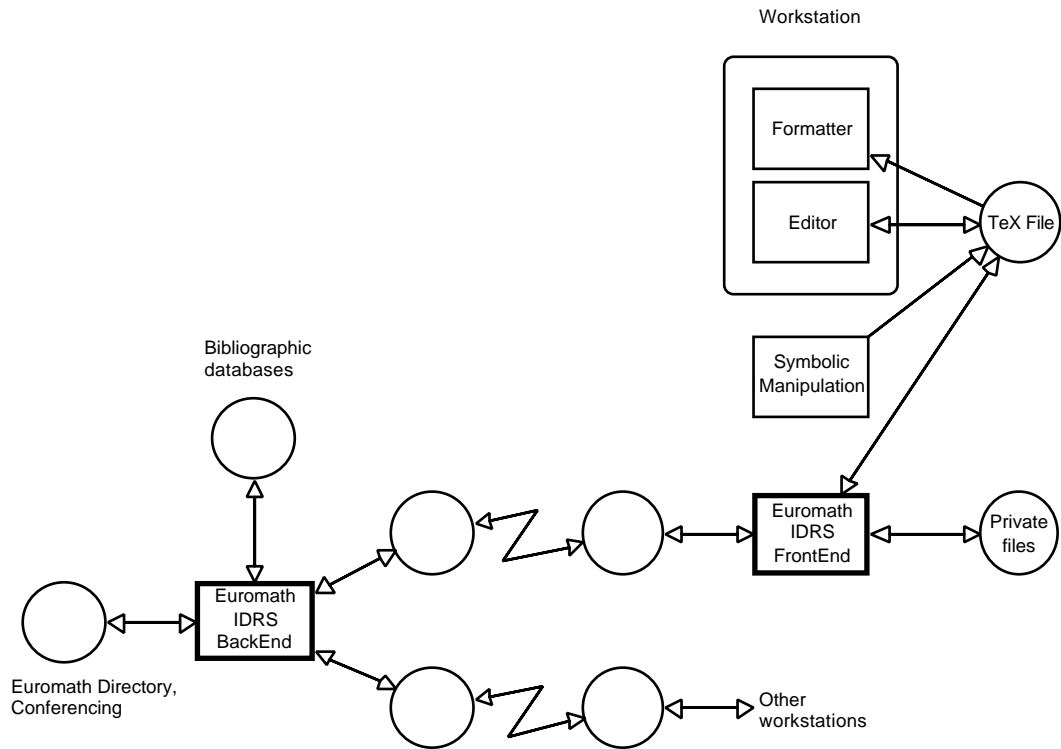


Fig. 9. A feasible EUROMATH architecture.

Literature

- Dennis Arnon, Richard Beach, Kevin McIsaac, and Carl Waldspurger: "CaminoReal: An interactive mathematical notebook", *Proc. EP '88 Conference, Nice, France, April 20-22, 1988*.
- Randall Collins and Sal Restivo: "Robber barons and politicians in mathematics: a conflict model of science", *Canadian Journal of Sociology*, vol. 8, no. 2, 1983, pp. 199-227.
- Lyne DaSylva: "First Field Study of Electronic Mail and Conferencing System", EUROMATH, Task 2.1.3.: Field Study of Interpersonal Communication, Dansk Datamatik Center, 1987.
- Philip J. Davis and Reuben Hersh: *The Mathematical Experience*, Houghton Mifflin Co., Boston 1981.
- J. D. Gould et al. "Why reading was slower from CRT displays than from paper", *Proc. CHI+GI 87, Toronto*, pp. 7-11.
- European Mathematical Council, Database Committee: *EUROMATH: the European Data Base and Communication System in Mathematics*, April 1985.
- European Mathematical Council, Database Committee, Subcommittee on Normalization (M. Demazure, C. Mulvey, J. Valenca): "ESMC: European Standard for Mathematical Communication", Final report, 12 January 1986.
- European Mathematical Council, Database Committee, Functionality Subcommittee (M. Hazewinkel, A. K. Seda, B. Wegner): "Report of the Functionality Subcommittee of the EMC Database Committee", 22 January 1986.
- Charles S. Fisher: "Some Social Characteristics of Mathematicians and Their Work", *American Journal of Sociology*, vol. 78, no. 1, 1973, pp. 1094-1117.
- J. Sutherland Frame: "The Working Environment of Today's Mathematician", in T. L. Saaty and F. J. Weyl (eds.): *The Spirit and Uses of the Mathematical Sciences*, New York, 1969, pp. 48-67.
- Richard Goldstein, James Loomis, and Avram Tetewsky: "Technical Wordprocessors for the IBM PC and Compatibles, Report by the Boston Computer Society. Part I TWP Capabilities and People Needs", *Notices of the American Mathematical Society*, vol. 34, no. 1, January 1987, pp. 15-32.
- Warren O. Hagstrom: *The Scientific Community*, Basic Books, New York and London, 1965.
- J. L. Hadamard: *The Psychology of Invention in the Mathematical Field* (1945, 2nd ed., 1949); Dover, 1954.
- Michael Hallett: "Towards a Theory of Mathematical Research Programs, I", *British Journal for the Philosophy of Science*, vol. 30, no. 1, 1979, pp. 1-25.
- Michael Hallett: "Towards a Theory of Mathematical Research Programs, II", *British Journal for the Philosophy of Science*, vol. 30, no. 2, 1979, pp. 135-159.
- Michiel Hazewinkel: "Technical Word Processing and EUROMATH", *Wheels for the Mind Europe*, no. 2, August 1987, pp. 15-20.
- Ernest Nagel and James R. Newman: *Gödel's Proof*, Routledge & Kegan Paul, London, 1959.
- Niels Vigand Pedersen and Flemming Topsøe: *Working paper on the plans to create a European Data Base and Communication System in mathematics*, Danish Mathematical Society, October 31, 1984.
- Richard S. Palais: "Mathematical Text Processing", *Notices of the American Mathematical Society*, vol. 33, no. 5, October 1986, pp. 741-751.
- Thomas Schøtt: *Fundamental Research in Small Countries. Mathematics in Denmark, 1928-1977*, Matematisk Institut, Aarhus Universitet, Ny Munkegade, DK-8000, Aarhus C, Denmark, 1979.
- Richard Startup: *The University Teacher and his World. A Sociological and Educational Study* (1979), Gower Publishing Co., Aldershot, England, 1981.
- Richard Startup: "The Sociology of Mathematics", *Sociology and Social Research*, vol. 64, no. 2, 1980, pp. 151-167.
- Raymond L. Wilder: *Evolution of Mathematical Concepts* (1968), Transworld Publishers, London 1974.