Book Reviews

The Book Reviews section is intended to cover a wide range of literature related to history and computing and related subjects. Review articles as well as reviews of individual works will appear.
Material for review and items of correspondence relating to reviews should be sent to the Review Editor, Dr Peter Denley, Department of History, Queen Mary and Westfield College, University of London, Mile End Road, London E1 4NS. e-mail: p.r.denley @ qmw.ac.uk

Evan Mawdsley and Thomas Munck, *Computing for Historians: an Introductory Guide*. Manchester University Press, Manchester, 1993. xvi + 231 pp. ISBN 0–7190–3548–1 (pb). £10.99 (pb).

In somewhat simplified terms, I would suggest that there are two groups that authors of an introduction to history and computing might wish to reach: first, the sceptics who doubt the usefulness of computer applications (or are downright suspicious) but who feel they are being pushed towards demonstrating the use of IT skills in their research or who feel required to show some understanding; second, the willing converts who, for a variety of reasons, would like to explore the use of computer techniques in their work. The authors of this introductory guide to history and computing are therefore confronted with the difficult task of providing a readable, accessible, introductory guide written for an intelligent audience who may require concrete evidence of the applicability of computer techniques in historical research.

The difficulty with many introductory guides is that they tend to be written by those with many years' experience and a great deal of knowledge of the subject to which they are introducing the reader. Unfortunately this can lead to a significant gap between what the authors regard as an introductory text and the information that a novice would regard as introductory. I am sure that most readers have experienced that overwhelming sense of disappointment on picking up a book (or even worse, of buying a book) which claims to be an introduction to the subject one is trying to come grips with, only to find that it is an 'introduction' which assumes a fairly sophisticated level of understanding of the subject and of the jargon.

It is reassuring, therefore, to read the preface of Mawdsley and Munck which very clearly specifies the purpose of the book and the target audience;

> the book is designed above all to be understandable to those who have never used a computer before; it is about 'entry-level' historical computing. We believe that there are many historians who sense that computing is somehow 'for them' but are unable to get background material to hasten their progress. We aim to provide that background. (Preface, p. xii)

It is equally reassuring to read on page 6 of the introduction;

> the computer does not need to create a new type of historian or a new methodology; it is a tool suitable for all historians. (p. 6)

The book builds neatly in a clear, linear fashion, moving the reader through the complexities of computer systems and wordprocessing; into a case study using census material; onto collective biography; flat files; relational files; tables and spreadsheets; numerical data; record linkage; using existing databases; text analysis; hypertext; CAL; until finally the reader comes to rest in the highly informative glossary which de-mystifies the jargon and the bibliography which offers further delights for the, by now, semi-computer-literate historian.

The case study is extremely well presented. It demonstrates how to create, modify and analyse a database from an historical source. The four chapters direct the reader down a structured, easy to understand path addressing the issues surrounding data entry; structuring the source in a database; basic coding and standardisation; running basic queries; enriching the data by more advanced structuring and coding; and finally, some simple statistical analysis — description, correlation and sampling.

In addition, the authors offer the reader the opportunity to go one step further. In an appendix they provide a typed transcript of census records for 1851 for 100 people from the Gorbals district of Glasgow. This enables the reader to create their own database, either as

a paper and pen exercise or, preferably, as the 'real thing' by typing some or all of the records into a computer. This allows the reader to follow the case study with 'hands-on' experience — surely one of the most effective ways to learn — and all in the privacy of one's own home or office where one can swear, shout and make a fool of oneself with impunity — something this reader much appreciates.

The authors make good use of examples from 'real' research illustrating the techniques, methodology and the range of analyses that a relative newcomer could successfully carry out. For example, the chapter on collective biography (prosopography) is discussed using a database of the French parliamentary Convention of 1792–95; chapter nine looks at 'tables and spreadsheets using voting records and the census' and chapter eleven looks at 'linking sources: comprehensive analysis of a community' using a database of material from Odense in Denmark. By illustrating computing techniques using worked examples the authors keep the discussion clearly within the historical discipline. Indeed, the discussion throughout the book is firmly rooted in the historical with the computer regarded as a tool which enables the historian to develop and extend her/his analysis.

The book might have been improved slightly if less emphasis had been placed on the second and third chapters which deal with computer systems and wordprocessing, if more information on network services and how to access them had been provided, and if the authors had covered in greater depth the crucially important issue of documenting the data creation process. As the book stands this subject only warrants passing reference. It might have added to the historical interest to see a greater temporal spread in the sources used, perhaps demonstrating the use of sources from the early modern and medieval periods. It should also be noted that ICPSR do not accept enquiries from individual academics. Enquiries and orders for data materials from ICPSR should be directed through the ESRC Data Archive as the UK representative.

I am also slightly uneasy with the emphasis that is placed upon the database approach to historical computing. It is not until the book gets to chapter twelve that it gets 'beyond the database' to look at text analysis and hypertext and then relatively little space is given to examining how they might be used in historical research. The book gives the impression that these alternatives are more complex and difficult for the novice and perhaps not to be considered until competence in other computing techniques is achieved. It may well be the case that these techniques do present a range of problems for the beginner to consider but it would be a great shame for the historical community to slip into becoming database junkies in much the same way as some in the social sciences became SPSS, rectangular file junkies — cold turkey, so I am told, is not a pleasant experience.

Historians need to be encouraged to learn a variety of computing methodologies and techniques and to select the methodology most appropriate to the source or sources to be used and the questions to be asked. Technological developments are such that moving data between applications is becoming increasingly easy and newcomers to history and computing should be encouraged to be creative in their use of the computer. However, perhaps it is the case that this criticism highlights the tension inherent in all introductory books — how far does one delve into the more complicated issues and techniques before completely confusing the novice reader and no longer producing an 'introductory' book?

Throughout the book the authors have emphasised the use of historical skills and how they need to be applied when using a computer. They have stressed the need to engage with the source material and the questions to be asked of the sources, and that much of the work goes on away from the keyboard. Placing the use of computers firmly in an historical context should serve to reassure the sceptic that using a computer does not make you less of an historian, and to convince the willing newcomer that applying traditional historical skills to computer techniques is not too frightening a proposition. The book is a genuine introductory guide to history and computing and well worth the £10.99 asked for the paperback edition. Recommend it to sceptics, hopeful converts and historians at all levels — this is one book that won't disappoint when picked up off the shelf.

Sheila Anderson
ESRC Data Archive, Essex

Storia e multimedia. Atti del Settimo Congresso Internazionale/ Proceedings of the Seventh International Congress Association for History & Computing, Bologna 1992, a cura di/edited by Francesca Bocchi & Peter Denley. Bologna: Grafis Edizioni, 1994. pp. xix + 861. ISBN 88–8081–000–6 (pb), Lit. 70.000 (£28.00).

'Seven fat years. Contrary to biblical precedent, we see no signs whatsoever that they will be followed by seven lean ones; on the contrary, historical computing is clearly going from strength to strength.' Thus Francesca Bocchi and Peter Denley wittily summarised the remarkable results achieved by the Seventh International Congress of the Association for History and Computing. The large number of participants and the great success of the meeting, until then unprecedented in the history of the association, wholly justified their biblical analogy. The

91 articles and almost 900 pages of the Proceedings, now published under their editorial supervision, should easily convince anyone who was not in Bologna that their remarks hold true.

The Proceedings are a faithful reflection of the Conference, both in their structure — there are three main areas, Historical Research and New Structures for Historiography, Methodology and Educational Technologies, divided into several groups and subgroups giving 23 sections in all — and in the style of the contributions, most of which retain at least an echo of their original form of delivery. Although some Italian speakers presented their papers in English, 47 contributions are written in Italian, 39 are in English and only 5 are in French. This linguistic imbalance is obviously a direct consequence of the meeting's location, but is not to be deplored. The seventh Conference of the AHC also provided the opportunity for a detailed snapshot of the discipline's status in Italy. What some will regret, though, is that summaries in English have not been added to what amounts to more than half of the book. The e-mail address of the authors, whenever available, would also have been useful. There is no comprehensive Bibliography, but an Index of Names and a separate Index of Software close the volume. They are a useful facility, and one that I frequently miss in similar publications, but they both need to be used with the usual *granum salis*. They are not complete nor always precise, and in a few places they suffer from the same degree of typographical inaccuracy one can spot in other parts of the text.

On account of the enormous variety of topics covered, the different techniques and software employed by so many historians, and the various degrees of historical significance that each project actually succeeds in gaining before the reader's eyes, even a cursory sketch of all the contents of the volume would be an impossible task for a reasonably short review. What may be attempted instead is a presentation of some common features that emerge from the volume as a whole. As usual with such collections, the reader is invited to skip through the book rather than study each page. But those who will have the perseverance to read the entire volume will probably come away with an impression of overall harmony. It is a personal view of this unity that I now wish to offer to the reader, with a bias towards methodological issues rather than straight applications.

* * *

My first observation can be put in a nutshell: the title — Storia & Multimedia — is very well chosen indeed. In the papers given at the Conference there is a widespread tendency, sometimes more than implicit, to stress the passage from the second age of IT — represented by the mass diffusion of PCs and their user-friendly software so typical of the eighties — to the third age of multimedia, desktop publishing, CD-roms and networks, all features that have started to characterize and differentiate the nineties. Even in the limited context of word-processing alone, many authors have noticed radical changes, for we have moved from the PC as mere typewriter to the PC as typesetter (read what Giorgio Politi writes on the new characters available, and the increasingly cheap ways to reproduce them on paper), while OCR technology has proved vital for the digital conversion of entire corpora (V. Concepcion & D. D'Amato illustrate methods for assessing the results of an OCR process, and E. L. Helsper, L. R. Schomaker & H.-L. Teulings focus on OCSR, optical cursive script recognition systems). Interactive multimedia are becoming more and more important, not only for presentation of coffee-table products, but also for research, by providing modelling tools with open reconstruction capabilities (F. Colson, J. Colson, H. C. Davis & W. Hall illustrate Microcosm, a link-service environment for multimedia reconstructions; Celozzi Baldelli provides a more general discussion of the methodology of multimedia). And if the essence of the IT revolution consists in the digitalization of any form of knowledge, then recent years have witnessed the crossing of the third semiotic frontier represented by the visual domain of images (Michael Greenhalgh provides a defence of the analogue against the digital, and Giuliano Pancaldi sketches the possibilities opened by multimedia for the study of scientific iconography). Nor should we forget, in the context of such radical transformations, the enormous importance of the new model of 'spineless' textuality represented by hypertext, an instrument more and more widely used in all sort of applications, and whose logical structure has been under constant refinement for years now (A. Gisolfi, W. Balzano & A. Dattolo discuss hypertext models based on Petri-Net). Digital images, hypertext and user-friendly iconographic interfaces remind us of the renewed importance of visual thinking (Ramazan Acun underlines the importance of visual interface). And although the Internet fever is a phenomenon of the last couple of years, networks receive their due attention in some of the most far-sighted papers published in the book (L. Cajani & S. Lariccia outline a network for historical teaching).

Such a lively perception of a transition between two phases in the history of digital technology is not a mere matter of record. On the whole, there is a tendency to appreciate the new role of the historian as an expert system within the technological environment of Augmented Intelligence (the new, instrumental reading of AI), while in several papers the recent transformations in the field provide a motive for reflection on past experience. Hence, a second thread that runs through the

Proceedings is represented by the analysis of the development of some fundamental tools and their best applications during the last ten years. Attention is obviously centred on data management. Again, two fundamental issues emerge as constant themes in the methodological core of the discipline. Historical data-banks in prosopography (but not just in prosopography; see the article by Dino Buzzetti on a data-bank of masters and books in fourteenth-century Bologna, and the one by Giovanni Gasparotto on the study of the sources of Isidore of Seville) tend to be more source-oriented than in the past, for they must provide sufficient and reliable information for future comparative studies (Gunner Lind's article is a good starting point for further discussion; his flowcharts are interesting, if controversial). A look at Peter Denley's presentation of KLEIO — the source-oriented DBMS which makes possible an integrated treatment of text and images, devised by Manfred Thaller and developed by the Max-Planck-Institut für Geschichte at Göttingen — illustrates the two points of source-orientation and comparative method with all the necessary lucidity, although the reader may also wish to read Toine Schijvenaars' more critical article about the comparison between a source-oriented and a topic-oriented approach. On the whole, there is widespread agreement that the selection of information for a particular piece of research should, as far as possible, be kept separate from the original electronic recording of data, on which that research is based. But reservations, or even more radical doubts, emerge once such a methodological point is put to the test. How far should the preservation of original data go? An old historical question, which in some extreme versions of the source-oriented approach may become an insoluble paradox. The macro-context — understood as the semantic environment of the system of our significative elements — may amount to the entire world conceived as the class of all facts. More modestly, data-banks should probably aim at what in Soft Systems Methodology (SSM) is called a rich picture of the information to be managed. Any sort of data — linguistic, quantitative, structural, visual, topological and so forth — can be taken into account only in so far as the historian is aware of its relevance to the topic under investigation. If the historian does not want to run the risk of transforming his or her own project into an impossible quest for a global encyclopedia, the final collection cannot be altogether uninterpretative, and the role of notions such as critical choice and direct relevance must remain central.

A sense of radical transformations, a constructive evaluation of new developments in the light of past experience, but also a curious glimpse at future prospects: here is a third feature of these Proceedings. Some will argue that this is not an unusual disposition among computerised historians. From many authors comes the sensible request for better and more extensive standards, in all fields, from text encoding to database structures. Several projects present more or less home-made tools, developed to be better suited to the particular needs of the historian than the usual commercial packages can be. A lot more is expected to be forthcoming thanks to further, predictable improvements in various technological areas. There are requests, efforts and expectations which unite many authors but that are not unprecedented. So one may wonder where the novelty lies. My impression is that it is to be detected in a new attitude that permeates most of the papers, a sort of assertiveness which, I would contend, was not there in the past. For the issue now is no longer to see whether IT can be used in historical research and didactics and, if so, how far such technological applications can be expected to extend on the basis of whatever the engineers can make available to scholars. Things have moved much further, and now the point has been reached at which the need for applications in historical computing could start influencing the commercial development of new tools. Historians no longer look like passive buyers and users of market technology, but educated promoters of new, specialized applications. Until some years ago IT innovations were technology-driven; today they are more market-driven and in this respect historians can be a rather powerful force. Passive acquiescence to commercial packages belongs to the eighties, as some of the historians in Bologna seem to have perceived (I would suggest to read from this perspective Andrea Bozzi's article on software for philologists, Paola Moscati's panoramic view on computing and archaeology, M. Breschi & L. Pozzi's article on a package for demography, Giancarlo Palombini's article on ethnomusicology and what Leen Breure writes on XBase-software). Of course, the scattered suggestions to be found in these Proceedings are only a first step towards software and hardware developed with the specific needs of the discipline in mind, but further plans have been laid down in other contexts which go very much in the same direction, most notably the decision taken last October by the National Institute of Standards and Technology (NIST) to finance a five-year, $150m project to create a software component industry. It seems likely that in the not too distant future scholars will be able to buy components (standard and interchangeable parts) and fit them together to make their own software.

As may be already clear, various papers in this volume share a general awareness of the mature status reached by the sector. This is clearly discernible, for instance, in D. Erlach & M. Reisenleitner' s paper, which is more a synthetic and informative discussion of some central issues in data modelling and user interface design than a presentation of the less intriguing 'Prosopographical

Onscreen Manual of the Austro-Bohemian Nobility in the Early Modern Period' that the title suggests. A comparison with anything done in the field of Philosophy and Computing shows how much less revolutionary the use of IT in historical research and didactics has become nowadays. Indeed, this is so true that the discipline seems now to run the opposite risk of becoming a mere fashion — and not only among medieval or modern historians. Bologna has seen archivists, economists, historians of ideas, cartographers, philologists, art historians, demographers, historians of literature, geographers, and many other scholars and social scientists speak to each other as they probably have had very few occasions to do in other contexts. An interdisciplinary approach is thus a fourth feature that emerges with peculiar intensity from the pages of this volume. An extraordinarily wide spectrum of subjects and approaches has been brought together by a common methodology, and the merits of such a process of unification must be credited to the active members of the AHC.

So much for my view of the unity of *Storia & Multimedia*, a book that certainly provides a wealth of information and suggests a number of interesting ideas. If now I were asked to suggest another four areas that I would like to have seen covered at the Bolognese Conference, I would probably come up with something along the following lines:

1) a deeper concern for a culture of selection. Scholars have worked for too long in a context of shortage of data not to face now the danger of being overwhelmed by an unrestrained, and sometimes superfluous, profusion of information. Some form of scientific censorship will have to be developed and learned in the future, if we are to survive in a context of total access to the whole of human encyclopedia. It is to be hoped that the care exercised in avoiding loss of information in the process of input — which has prompted the debate on the source-oriented approach — will soon be parallelled by equally important care devoted to the selection of the results of the output process;

2) more substantial theorization on the use of quantitative analysis in intellectual history. Several of the papers involve quantitative techniques (two interesting examples are provided by Maria Pia Guermandi and Cornelius Steckner, while Dario Tomasella traces the roots of the approach), but there is still a lot of effort to be made in support of an interdisciplinary method of quantitative analysis in the field of intellectual history before a satisfactory foundation is laid for a mature ideometry;

3) applications of citation analysis. Sociologists of knowledge and historians of science have developed fruitful methods of research based on the number of citations received by a particular object of study (an author, a book, an article, a school, a topic, etc.). Also with an eye to a future ideometry, it would have been interesting if some space had been devoted to the use of citations in the history of ideas, and it is certainly to be hoped that quantitative historians will adopt such methods more extensively for some of their work in the future. After all, the study of citations is a clear recourse to what Jean-Philippe Genet has defined as a metasource, i.e. data that, though originally not structured in a quantitative pattern (the citations), become material for quantitative analysis through their collection in a data-bank;

4) a more vigorous effort, on the part of some projects, to rise from an initial stage of local chronicle to the level of a proper microhistorical reconstruction. The technology available has brought a welcome decentralization of the discipline, and one that is unprecedentedly radical. But I suspect that such a breakdown of disciplinary taboo, and the emergence of the periphery, should be followed more consistently by the firm acquisition of a sound micrology, a historical method of microanalysis capable of unearthing the universal in the particular and of preserving the unity of individual facts within wider perspectives. It is only on the basis of a "holographic" postulate, which invites us to discern in small details and apparently borderline events important fragments and traces of more general and all-embracing macrophenomena, that we may be able to justify our work on very limited topics that otherwise would remain of very marginal significance. A micrological, computerized approach to history should make it easier for us to approach the essence of human deeds and culture through analysis and comparison of phenomena on a reduced scale, rather than lead us to a short-sighted exaltation of minutiae and trivia. With the pace at which the discipline is progressing, the mere feasibility of a computerized project will soon no longer amount to a scientifically valid justification for its implementation. Global strategies are becoming vital.

* * *

The Conference was held in 1992. Since then many of the projects presented in Bologna have developed in more or less successful ways. To select just one important example, the possibility of an English version of KLEIO, envisaged at the time, is now a reality. Scholars with various degrees of involvement in historical computing have had two more chances to meet and compare their experiences at the annual Conferences organized by the AHC in Graz (1993) and in Nijmegen (1994). At the same time, technological changes in the field have been as radical and quick as ever. One has only to think of the new machines commonly available in the shop, with their double speed CD-roms and Pentium chips as stand-

ard features, or recall the recent diffusion of WWW browsers like Lynx and Mosaic. It is the endless dialectic of matter and change, stability and movement, which nowadays is being translated into that of velocity of microprocessors and quantity of bytes to be processed. As a result of such changes, some of the papers in the book are somewhat outdated, and the Proceedings themselves look more like a historical document than the latest source of information on the state of the art. I suppose one is entitled to regret that they were not collected at the time of the Conference and put on the Internet, or made available as an electronic book consisting of a floppy-disk and a brief guide. It would have been much cheaper, and the information would now be far more easily available than it is going to be. The volume is not listed in the on-line catalogue of Global Books in Print, and I expect one will be charged for mailing costs, if it has to be ordered from Italy. Moreover, the files could have been browsed, and anyone who wished to have a hard copy of an article would have been able to download it and print it. Enough. I do not need to argue for advantages which are probably all too well known to the reader. Perhaps the choice of printing the Proceedings has been forced upon the AHC for academic reasons. Electronic-verba volant, paper-scripta manent. If one needs to convince a Committee, a Foundation, an Academy or the University of the value of one's own research one must be able to provide evidence of one's own work on paper, like any other colleague. If this is the logic behind the publication of the Proceedings on paper, and I am only entertaining a suspicion, my suggestion would be that it is time scholars were more daring. Things change rapidly, and in some cases even the bureaucratic mentality may not take geological eras to update itself, especially if scholars were to decide to move firmly towards the electronic medium for some of their editorial work. From this perspective, it is comforting to see that steps in this direction have already been taken in connection with the last two Conferences. It has required quite a long time to made the Bologna Proceedings available in the shape of a book. Meanwhile 'the famine [of information] has been in the land [of knowledge] these two years'. Let us hope we shall not be kept hungry again in the future. IT does not allow for biblical time scales.

Luciano Floridi
Wolfson College, Oxford

Laboratoire d'Études et de Recherches sur l'Information et la Documentation, *Histoire et Informatique: Une bibliographie internationale 1993/ History and Computing: An International Bibliography 1993.* Halbgraue Reihe zur historischen Fachinformatik, A24, St. Katharinen, 1994. ISBN 3–928134–88–4. 170 pp. £7.50.

In the Foreword, Suzy Pasleau and Christine Puit explain the genesis of this bibliography beginning at the AHC Bologna conference in 1992. The aim is to create an annual bibliography. This volume not only covers the publication period of 1993 but also, in order to create a perspective in the subject, the years 1990–93.

The bibliography contains monographs, and articles in journals and periodicals. These 688 items are laid out according to a classification devised by LERIDOC. This is always a difficulty as an author does not always (or even sometimes) conveniently stick to one subject. Each bibliographical item consists of the bibliographical information together with an abstract, keywords, primary/secondary sources covered, whether the title contains a bibliography, tables, graphs or maps, the time period the contents cover, its geographical area, what software may be discussed. Not all of these 'fields', however, are always present. The original language of a title is retained, sometimes with a translation given in English or French.

There are three indexes: author, software/hardware/language (divided into three), and a list of the 23 journals and periodicals reviewed. The table of contents is given at the front with a French version at the rear as is conventional.

(Some) titles are assigned time-periods which is a useful device to indicate potential suitability to the reader. Most periods are in century form, e.g. 1800–1920, or 16e–18e siècles, the earliest period is described as Moyen-Âge. I have problems with this schema on two counts. First, one person's Middle Ages could include another's late antiquity. Secondly, why not stick to one system, such as, 1400–1500 or 15th–16th centuries, but please, not both! The dating is not always correct (see 94–571, the title contains the date 1140–1803 but with period Moyen-âge only). The century should be sufficient with the added benefit of not becoming complicated (as in 94–577, for a title covering 1700–1914 with period 16e–18e siècles, 1800–1920, another much simpler way surely could be: 18th–20th centuries?).

In that niggly way that reviewers have, I tried to imagine topics for which I as a user might use this bibliography as a source. As BRS/Search is a major piece

of software in my work, I looked this up in the software index: no entry. It does occur under Orbis, so puzzled I looked up the entry and deduced that BRS/Search is the software used in the Orbis project (made even more interesting to me as the Orbis I know is a module of NotaBene!).

Another topic of personal interest is OCR (which is covered in none of the indexes) and is surely an important subject for historians — how is one to find out about other people's experience of this? e.g. the contents of *Optical Character Recognition in the Historical Discipline* A18 (1993) (entry 94–675 in this bibliography). OCR is sometimes indicated in the keyword 'field' but is not accessible through the indexes.

It is easy to be picky but consistency in first names (M. Thaller and Manfred Thaller occur), software names (AskSam and ASK-SAM), etc. would be nice. Likewise keywords are in English or French, no doubt due to the differing languages of the contributors, but centralized editorial control, not only in this area, would greatly add to the general appearance.

In general, this is a brave attempt to begin gathering together titles such as these. Bibliographies are never easy to compile, never mind to present. Judging by the seminar session devoted to discussion of this Bibliography at the AHC Nijmegen conference, there is considerable interest in this undertaking. The problem with bibliographies is that like how families should be run everybody thinks they know how to do them — many hands make light work but too many cooks can spoil the broth! There are faults with this volume but it is also a good start and should be supported.

Margaret Lantry
Faculty of History
University of Cambridge

Optical Character Recognition in the Historical Discipline. Proceedings of an International Workgroup organized by: Netherlands Historical Data Archive, Nijmegen Institute for Cognition and Information. Halbgraue Reihe zur Historischen Fachinformatik, A18, St. Katharinen, 1993. ISBN 3–928134–97–3. 165 pp. £7.50.

Optical Character Recognition (OCR) is the great bottleneck of historical computing. The problem can be outlined briefly and simply. To carry out historical analysis sources need to be in machine-readable form. But how do we convert hard text into the digital form recognised by the internal architecture of computers? The most common answer to this question is that we enter the data manually through the keyboard. Or rather we get our poorly trained research assistants to do it, unless we've got enough money to pay a gaggle of poorly paid wage-slaves from the Third World. A less common solution is getting the stuff scanned and how deceptively simple it all seems; after all the majority of commercial scanning packages promise 99% accuracy. So just get a whacking great research grant, buy a Kurtzweil, get your source, train the scanner by running a couple of pages past it and away you go. Would that life were so simple, a sentiment which might, given wackier editors, have been the subtitle of this collection of papers.

Optical Character Recognition is a useful book for anyone interested in the potential, or lack of potential, of OCR but with insufficient knowledge or know-how to proceed. The fifteen essays give an eclectic overview of what historians have been attempting to do with OCR, the problems they have met and the solutions they are employing. The scope ranges from the highly technical (Concepion and D'Amato's ahistorical 'Symbol Correspondence for Integrating Multiple OCR Outputs') to the we-might-do-it-this-way-if-we-get-round-to-it (Ann McVeigh 'The Irish Database Project: a Case for OCR?'). In between we learn that a lot of people involved in OCR are using, or contemplating the use of, neural networks (do not even consider reading Andrew Senior's 'A Recurrent Network Approach to the Automatic Reading of Handwriting' unless you can understand Rumelhart *et al.*'s pioneering *Explorations in Parallel Distributed Processing*, Volume 1).

But if you really want to know about the pros and cons of OCR then read Mark Olsen's 'Scanning, Keyboarding, and Data Verification: Factors in Selecting Data Collection Technologies'. Olsen and his team have actually tested different OCR packages. Furthermore, they have compared the time and cost of using scanning technology compared to 'off-shore' (read: 'a typist pool in Jakarta'?) keyboard data entry. His findings are salutary to anyone considering the purchase of a scanner:

> There are many instances where OCR is not cost-effective, since the technology is still prone to making numerous errors. Claims of extremely high accuracy for small samples of scanned material do not ... translate to high accuracy for large projects.

And there's worse news in some of the other contributions. Handwriting is, of course, a no-go area, a fact reflected in all four papers on this subject. We've barely got past first-base here. If your source is old, smudged, poorly printed, or in a non-English language then you'd have to come up with a pretty good reason for the purchase of a scanner, although Borodkin *et al.* do show that there has been some progress in scanning Cyrillic characters ('Applications of OCR in Russian Historical

Sources ...'). The only good news seems to be for those of us who are using sources which, like census data, consists mainly of numbers organised in columns (McVeigh and Olsen). For those 'lucky' enough to work with structured sources, just remember the problems caused by the shifting structure and meaning of these texts over time and get that bottle of aspirin off the shelf again.

Despite the problems, or even because of the problems, the kind of research effort that this book represents is important and invaluable. Those of us who use computers tend to think we've entered an era of unparalleled sophistication. Thus the past history of computing is always represented as a savage and simple country: a Whiggish viewpoint reflected in the many comments throughout this collection which stress how far things have come. Of course this is a load of tosh. For the reader, on putting this volume down, is more likely to say wistfully: 'how far we've got to go'. Compared to what the future holds we're just kicking pebbles around the seashore. Or rather, the contributors to this work are kicking the pebbles (I'm staying well clear) and a good job they seem to be making of it. It is through such scholastic effort that the big breakthroughs might eventually come. For history, with its diverse and complex sources, provides the extremes with which scanning will have to cope if it is to be a truly intelligent and useful technology.

James Bradley
Wellcome Unit for the History of Medicine
University of Glasgow

Beatrice Moring, *Skärgårdsbor: Hushåll, familj och demografi i finländsk kustbygd på 1600–, 1700– och 1800–talen.* (Household, family and demography in Finnish coastal regions during the 17th, 18th and 19th centuries.) Helsinki, 1994.

This book challenges two major theories about the history of the family in Western Europe. For one thing an extended family pattern dominated the area before the 19th century. Second, the mean age at marriage is found to vary over the centuries.

The Åland archipelago is situated in the Baltic sea between Stockholm and Finnish Åbo. Today it has a semi-independent status under Finnish sovereignty, but the islands belonged to Sweden until 1808 and the population is ethnically Swedish. Moring has studied the demography and household structure of this society, focusing on economic development, marriage patterns, social and geographic mobility, fertility, illegitimacy and mortality. She finds a 17th and 18th century society dominated by peasants who by law monopolized the fisheries. Cotters were not allowed to establish themselves. For these reasons, most households were complex with multi-generational cohabitation; for instance grown-up siblings could form a household. The few proletarian families had a more nuclear structure. From the middle of the 18th century this became more common among the peasants as well. At this time the law against cotters was abolished, and the peasants lost their monopoly to the fishing grounds. Other reasons behind this development were new fishing techniques that made it easier for smaller and poorer groups (even individuals) to make a living from fishing alone (not combined with farming). As a corollary to the changing household structure, the mean age at marriage rose from about 22 years around the beginning of the 18th century to 27 years in the 1820's. The proletarian marriage age had all the time been higher than that of the farmers.

Circular migration exchanging marriage partners between the islands dominated the geographic mobility pattern, while long distance out-migration over time changed from predominantly Swedish to Finnish destinations. In-migrants were scarce, but during the 19th century people from hunger-stricken Eastern Bothnia settled on some of the marginal islands.

The author deserves praise for having followed her population of study through more than two centuries. To achieve this she has used two distinct techniques. The analysis of family structure is based on cross-sectional methods, using the census-like 'mantalslängder' for the working part of the population and catechismal church records for the children. Even if these sources do not cover the entire population, the cross-sectional methods deal with several municipalities in the archipelago, so we can hardly raise questions of representability against the results on family structure.

For the more demographic issues a family reconstitution study was undertaken, covering families with children born from 1742 to 1825. To save work, however, this had to be limited to the single parish of Houtskär. There are 547 reconstituted families plus 63 that could not be completed due to migration. Out of these, 350 comply with all the demands in a family reconstitution study. One must, therefore, ask to what extent the reconstituted families can represent the whole Houtskär population. Even more serious is the question about the representativity of Houtskär in the whole archipelago. Moring claims her results to be representative, but on page 30 she admits that Houtskär had less farm land per inhabitant than the other communities she investigates. Thus, for instance the finding that downward social mobility hit one third of Houtskär's population, is not automatically valid for the rest of the islands. Another reason one should be wary about generalizations is the small numbers in many tabulations where separate group-

ings are analyzed.

Those who want to retrace the steps of Moring's research will have much work to do, since the contents of the source material was standardized before computerization. This procedure is maybe more effective in the first round, but checking and secondary analysis becomes more difficult.

A reviewer is always tempted to ask for more than the author was prepared to offer. Particularly I miss a discussion of to what extent there was a time lag between the amendment of the laws and introduction of new fishing techniques on the one hand and the changes in household structure on the other. Also, a more systematic comparison of Åland with Swedish and Finnish findings would have been useful. Fortunately, Moring's study is part of the joint Nordic research project 'Changing coastal communities 1650–1950'. Thus, interesting findings from all the Nordic countries are presented in a 1993 special issue of *Journal of Family History* (volume 18, number 4).

The critical remarks notwithstanding, Moring has written an important book that can definitely not be refuted as a special case from North-Eastern Europe. Let me conclude by citing her punch line: 'The dominant role of the nuclear family in English history may be a matter of culture, but also a proxy for the degree of proletarization. In few other places did the proportion of paid workers grow so high at an early point in time.'

Gunnar Thorvaldsen
University of Tromsø