

Enhancing and Diversifying ATMA's Databases to Ensure Sustainability

SUPYAN HUSSIN
DING CHOO MING

ABSTRAK

Ekoran daripada peningkatan minat dalam pengajian Dunia Melayu di seluruh dunia, lebih-lebih lagi perubahan dalam kepengarahan ATMA pada April 1999, kami di ATMA telah diberi peluang untuk pertama kali dalam sejarah UKM untuk membina pangkalan data sendiri. Setelah membina pangkalan data yang pertama, PADAT, kami telah membangunkan pangkalan data pelengkap yang lain, dari masa ke masa. Kesemua itu bukan sahaja untuk memenuhi kehendak, tetapi juga mengasah selera yang baru. Kejayaan kami setakat ini adalah hasil daripada pasukan kerja yang cekap, kepimpinan yang berwawasan jauh dan sokongan padu daripada pengurusan atasan UKM. Kami ingin membangunkan pangkalan data yang bertaraf dunia dan dalam berbuat demikian, untuk membuat ATMA menjadi satu lagi titik penentuan dalam pengajian Dunia Melayu. Untuk membuat pangkalan data itu dapat menyarai diri, kami mesti dapat menyakinkan pengguna yang membayar yuran nanti bahawa wang yang mereka belanja itu boleh mendatangkan manfaat. Memikirkan penyelidik kini sudah biasa menggunakan komputer yang lebih laju, selain menjadi semakin tidak bersabar, maka tekanan ke atas pangkalan data sudahlah meningkat. Kami mesti memastikan maklumat yang tepat boleh diakses dengan cepat pada bila-bila masa.

Kata kunci: IRPA, enjin gelintar, Visual Basic, Padat, Pantun baba, peribahasa Melayu, karya Jawi, kamus Melayu, pangkalan data pelengkap, dunia Melayu

ABSTRACT

Following the increased interest in the Malay World studies world wide, and a change in ATMA's directorship in April 1999, we had, for the first time in UKM history, the chance to design our databases. After creating the first database, PADAT, we have successfully created new supplementary or complimentary databases. They are constructed not only to satisfy demands, but also to whet new appetite. Our achievements thus far are the result of an efficient working team, a visionary leadership and overwhelming support from UKM's top management. We hope to develop our databases to meet world standards, and in doing so, to make ATMA to become another defining point in the study of the

Malay world. In order to create a circle of sustainability, we must in turn convince paying users that their expenditure will benefit them. As researchers get used to faster computers and become more impatient, the pressures on databases increases. The right information must be quickly accessible and at any given time.

Key words: IRPA, search engine, Visual Basic, Padat, Pantun baba, Malay proverbs, Jawi works, Malay dictionaries, complimentary databases, Malay world

DATABASES IN THE DIGITAL AGE

The proliferation of digital resources has brought dramatic changes to the scholarly world since the 1990s. With it, come different perspectives on how research collection, including databases, should be developed, managed and preserved. Furthermore, digitizing has now eased custodial and reference concerns and brought the collections directly to the users without staff intermediary. This is cost saving, fast and convenient for everyone involved. In response to this change in knowledge environment, ATMA started in 1999 to employ IT to enhance research and information management. Our decision to make maximal use of IT was due more to the increasing availability of the Internet and the decreasing cost of accessing it than to the falling prices of computer and communication technology. The capacity for handling large numbers of users simultaneously, and the reliability of reproduction and transmission at great distances were also important factors. Fully conscious of the power of digital technology in revolutionizing the way we capture, store, preserve, disseminate and cope with the exponential growth of recorded knowledge, we developed a blueprint to guide the development of databases at ATMA in our previous paper (Ding & Supyan 2002b). In that paper, we addressed the problems involved in collecting, extracting, repackaging and regrouping disparate works related to Malay World studies to support access by ATMA researchers. In attempting to turn vision into reality, we faced the complex problem of staffing, financing and equipment shortage. With a small grant from IRPA, we proceeded to extract and repackage articles from institutional and personal libraries, and designed a database with a search engine using Visual Basic. Until then, there was no locally generated database on Malay World studies anywhere in the Malay world. With the help of modern IT, we hope to transform scholarly research and communication in Malay World studies. Three years down the road, now with some 30,000 articles on various aspects on the Malay World studies assembled in *PADAT (Pangkalan Data Alam Melayu)* we feel confident enough to construct more databases related to the subject, to remedy insufficiencies, to stimulate new research interests and to ensure the sustainability of the databases.

Following the increased interest in Malay World studies world wide, and a change in ATMA's directorship in April 1999, we had, for the first time in UKM history, the chance to design our database, and to change the whole knowledge environment in a fundamental way! As a team, we find ourselves very lucky in the sense that we firstly are researchers, and secondly, two of the team members are IT experts and the third one an experienced information scientist. We have had the opportunity to involve our researchers and to understand their preferences and expectations. The building of the database itself was the most engaging project, fuelled by our enthusiasm and by rapid advances in digital technology. In order to accommodate increasing demands from researchers, we undertook strategic steps occasionally to expand *PADAT* and create new supplementary or complimentary databases. Here the Malay proverb "*sedikit-sedikit lama-lama menjadi bukit*" is clearly illustrated not only in *PADAT*, but also through the growth of other databases, namely the *Pantun Baba* (Ding 2002d), *Peribahasa Melayu* (Ding & Arba'eyah 2002e), *Jawi works and Malay Dictionaries*. All these efforts aim at broadening and deepening research at ATMA, and complimenting that which has already been done elsewhere, for example at KITLV in Leiden, ISEAS in Singapore, SOAS in London and CSEAS in Kyoto.

GROWTH, DIVERSITY AND SUSTAINABILITY

In the emerging knowledge environment of advanced IT, computers and networking, our clients may very well come to expect more than we can deliver. How will we then manage it? How can we accommodate both current and future users? How will they access information? Is *PADAT* alone able to cope with all their needs within the vast and multidisciplinary subject of Malay World studies? Can we consider all their goals and meet all their needs? How can we improve access to digital information? A trove of digital documents is not any better than having paper documents if instant retrieval is not possible. In order to meet researcher needs, we have to improve the database that we have and introduce new ones from time to time, carefully considering our indexing plans to accommodate their needs.

Our experience has grown and our options have increased over the years. We have managed well so far. ATMA's databases have now earned a reputation for addressing key issues in information management and retrieval in Malay World studies. *PADAT* has been improved and new databases introduced. These innovations have been made with at least the following objectives in mind:

1. To sustain the interest of current users.
2. To attract new customers.
3. To stimulate new interests.
4. To expand research opportunities.

5. To generate income.
6. To ensure the sustainability of databases.
7. To increase specialisation.
8. To handle the upsurge of diverse interests.
9. To search for new alliances and partnership.
10. To incorporate materials relevant to related fields of research and which meet the needs of researchers.

In other words, databases at ATMA are constructed not only to satisfy demands, but also to whet the appetite of old researchers and attract new ones with interests related to Malay World studies. Thus, the dramatic quantitative and qualitative expansion of our databases has taken place naturally, coinciding with the multidisciplinary research promoted, encouraged and conducted at ATMA. Over the years, research interests at ATMA have expanded in line with the overall interest in the Malay World. Nevertheless, the most important factor contributing to the rapid diversification of interests at ATMA in recent years is the deepening and broadening of multidisciplinary research, with new research adding new fields beyond Malay language, literature and culture. The development of these databases can be regarded as a natural move to maintain ATMA's position as a premier research institute.

In business, the route to rapid growth and big money is diversification, which is best achieved through acquisition, which reduces competition. Thus, we are bombarded daily with news about mega-mergers among big corporations, including those within the telecommunications and banking industries. Seen this way, diversification can be a strategic step taken by organisations which are not performing well, and which may even have lost sight of their original activities. Thus, according to Detrie & Ramanantosa (1986: 32), the stronger the competitive position an organisation has, the less necessary it is for it to diversify. Unlike the business world where diversification involves radical changes in the character of the business, our diversification does not occur by means of acquisition, but by introducing new databases in order to achieve a substantial impact on as many users as possible. They are also expected to hold as much appeal as possible since all of them have their own worth in relation to different researchers at different times. Archival and rare collections like *Malay Dictionaries by non-Malays* and *Pantun Baba Malaysia*, for example, have hidden and undiscovered value. They are gems that hold a potential for opening new avenues for research. Issues addressed in the creating and diversifying of ATMA databases are:

1. Broad research value,
2. Local research value,
3. High demand materials,
4. Unique materials,
5. The significance of a particular collection in relation to other collections.

These considerations fit into ATMA's medium and long-term development. In order to ensure that our users are able to find material to further their research, ultimately making them stay with us and financing us in our further development, we have to put their interests first. In other words, ATMA's strategy for diversification has been spontaneously fuelled by the needs of researchers. This will be discussed later. Now that more databases are making ATMA visible to more people, it is vital that existing databases protect and promote our competitive position. Detrie & Ramanantosa called such strategies "investment", "branching out", "support" and "survival diversifications" (1986: 35). As in other businesses, we believe that the stronger our position, the greater our chances to grow and the wider our opportunities for diversification. The better the starting position, the less risk will be encountered. Since research on the Malay World is our core business, we have to think about how best to feed our users' insatiable appetite for information in this specific area. We have not only to offer maximum benefits in terms of what can and should be done and when and with whom, but also to provide ourselves with a plan whereby we can make ourselves self-sustaining in the future. Beyond that, we must also consider our ability to develop a document delivery system that is responsive to our clients' needs. Indeed, besides the technological configuration of our choices, a significant issue that the new technology and the new scholarship have raised is "what to collect".

Practically everyone doing research on the Malay World will find the materials in *PADAT* to be of interest. However, the collection alone cannot possibly satisfy everyone's needs, and we seek increasingly diverse materials presently not available in *PADAT*. Since the database was designed specifically for collecting critical articles only, we have introduced new databases, including those for *Pantun Baba*, *Jawi works*, *Malay Proverbs*, *N. A. Halim's Collection* and *Malay Dictionaries compiled by non-Malays*. Despite their essential differences, they compliment one another in providing better service to client scholars, and in attracting new researchers. As far as sustainability is concerned, diversification and specialisation are complimentary and not conflicting phenomena. Given the astonishing array of old and new materials about the Malay World available in different media, formats and languages on the one hand, and the complexity of research needs, which often defy neat categorisations on the other, it was no longer advisable to maintain only one database. We cannot achieve the impossible, but that gives us all the more reason not to fail in doing what is possible! And since we cannot do everything, we must be sure that what we do is the correct thing. Given our limited resources in terms of finance, staff and equipment, we must make sure that any investment we make also contributes to the sustainability of ATMA's databases. To us, diversification is also about market penetration and geographical expansion, which are part of our medium and long-term development plans. Furthermore, there are a range of valuable collections that cannot be integrated into *PADAT* in any satisfactory fashion, such as our

Pantun Baba and Malay Proverbs. More importantly, diversification also increases the need for comprehensive yet specific classification. Thus, when one logs into <http://www.atma.ukm.my> or www.malaycivilization.com, one must also be able to gain easy access to other special collections at the same time.

PROBLEMS AND OPPORTUNITIES

It seems that ATMA's databases come about at the right time, and have been sprouting like bamboo shoots, almost with a dynamic of their own. Now, after more than three years developing databases, we need to think of our operations as a business and to look for ways to turn them into sources of steady income. Though the size of our business must be limited initially, we have to consider long-term profitability, growth potential and sustainability. Sustainability is very much a financial issue. It is easy to envision, and stimulating to discuss new projects, but it is difficult to find financing for them. In this regard, our budget reins our rhetoric and reveals our real problem. The meagre funding ATMA receives from UKM is insufficient to pay for staffing and ongoing digitising, regular copying, refreshing and migrating data, and upgrading hardware and software. We must therefore generate substantial income from other quarters. The occasional corporate or foundation grant, however generously given and gratefully accepted, will not get us to where we believe ATMA needs to go technologically – onto a par with other established research institutes. Since www.malaycivilization.com will be a permanent feature at ATMA, it is vital to secure annual funding for maintain the site. We cannot rely on the sporadic funding cycles, and must therefore develop a program of systematic funding capable of sustaining our database developing efforts and of meeting other institutional (*i.e.* ATMA) priorities. We have to be as creative in finding money as we have been in initiating projects. The stages involved are not easily carried out and the costs are not low. A program must thus be developed that can attract funding from external sources or justify the re-directing of resources within an internal budget already stretched thin.

Free service is but a myth! Huge sums are needed for hardware, software, Internet connections, and staff to do data entry and updating, to mention a few. Can we relay our costs to the end-user? After all, non-paying customers do receive services that are costly to provide. So far, we have not put a price tag on our service, relying only on our grants to get by. Some thought must be put into the suggestion that we start charging customers, be they individual users, schools, libraries, research institutes or media. Alternative payment systems include through annual subscriptions, *pay-per-use* or *pay-per-view*. We may also have to transcend traditional funding, and start attracting corporate funding. This may lead to the establishment of partnerships with the institutes concerned. The collaboration may also reinforce institutional commitment. At

present, we are working with *TM Net* and *Utusan Online* to attract as many users as possible to our products and services. We need to be sure that the potential user community for our databases is large enough to justify the costs.

In order to create a circle of sustainability, we must in turn convince paying users that their expenditure will benefit them. User needs and demands are complex things. They are subjective in that they include considerations such as reliability, speed of response, hours of access, level of detail, ease of use, range of choices and level of sophistication. Speed and convenience are vital to the existence of databases. As researchers get used to faster computers and become more and more impatient, the pressure on databases increases. The right information must be quickly accessible and at any given time. This fact cannot be ignored. We have to give ample consideration to the researcher's working conditions when we design retrieval methods. To keep users interested, we have to maintain visibility even as we offer information in new and flexible ways. Database access must be comprehensive. We must assume that our user group does change, which means that we must obtain up-to-date information about our customers. This is not an easy task in itself. Nevertheless, learning their preferences and expectations suggests new service opportunities. We need to master this skill. Saving our users hours of laborious searching, and increasing their productivity will make them more willing to remain paying customers. *PADAT* needs to develop into a world-class database for Malay World Studies, comparable to most other commercial online databases, such as *LEXIS* and *Westlaw* in the field of legal information.

All databases possess a variety of search procedures, ranging from Boolean searching to truncation that allows natural language queries. Here, word search is of paramount importance. So far, *PADAT* has worked very well as long as the single words or phrases used are relatively unique. Problems arise when relatively common words or phrases are sought, giving a avalanche of hits that are totally irrelevant. In such cases, users are encouraged to use Advance Search, which enables them to combine name of author, title of article, keyword and source. Here, Boolean logic allows them to build a query that contains many terms used in combination, while in wild cards, the query for '*fill*', for example, will retrieve all terms using the root word "*fill*", including such words as "*refill*" and "*filling*". Further improvements to the system will enable search by authors, titles, keywords, sources, accessions numbers, and will also arrange the hits alphabetically, chronologically or by accession numbers in ascending or descending order. Following the example of advanced databases, we also allow users a bird's-eye-view of the entire collection of articles on a specific topic, sorted by author, title or keyword. At the moment, more than one hundred people visit and use our databases daily. More and more of them have to their satisfaction managed to gain instantaneous access to relevant texts. Due to the issue of copyright, most articles in *PADAT* have yet to be digitised and provided with images, as has happened with our other databases. This is the major handi-

cap in *PADAT*. Nevertheless, all hits in *PADAT* are supported by their relevant documents available for physical reference at *ATMA*.

Diversification, or multiplication, appears to be an inevitable aspect of our drive to become competitive. We seek diversification while still in a strong position. Given limited resources, the creating and managing of databases are always about choices. We search for rare, important and archival materials when creating supplementary or complimentary databases. Co-ordination in a co-operative environment is a major concern, since these databases must relate in a comprehensive, and not scattered, manner, presenting a clearly defined and unifying purpose, as recommended by Matthews (1998: 194). The site www.malaycivilization.com will thus be composed of a web of many small and specific but related databases. The growing number of databases must also be supported by sufficient free random access memory (RAM), with the fastest possible connections to large capacity storage. There are four major issues where sustainability is concerned:

- i. System functionality over time: We must maintain the functionality of the systems and upgrade their capabilities as technology evolves.
- ii. Storage media deterioration and migration of data: It is necessary to move digital data and the accompanying index information to newer generations of imaging systems before deterioration sets in. Although we may not have exact information about media longevity at the moment, WORM (write once read many) technology has so far offered a greater level of data security. Besides that, we must ensure that backup procedures are a matter of routine.
- iii. Digital image data quality: Images should be captured at the highest possible resolution: 300 dpi should be the minimum, while 600 dpi may be necessary to guarantee fidelity to the original.
- iv. Integrity of information resources: The digitally converted materials are simply a random and unintelligible collection of files unless properly indexed.

In order for searches to be faster and easier, all the databases will be integrated under our portal www.malaycivilization.com. This portal was created with a grant from MOSTE (Ministry of Science, Technology and Environment) under DAGS (Demonstrator Application Grant Scheme). By linking all the databases, we give them increased visibility and status, satisfying our aim that the portal must be broad-based and merit-oriented.

CONCLUSION

ATMA is the first research institute on Malay World Studies to have interactive databases accessible on the Internet. They are the products of a bold step taken

by ATMA's 9th director, Prof Shamsul Amri Baharuddin, a man of vision and action. Now, database building and expansion are ATMA's defining commitment that not only aspire to fulfil the needs of researchers and scholars, but is also the thrust of ATMA's general plan for growth in the foreseeable future. In our endeavour to promote Malay World Studies, we combine the power of emerging technologies with important trends in scholarship. Technologies that we now associate with digitisation – www, databases and search engines – have made images, sound and full-motion video, etc., available to anyone with a desktop computer of a reasonable standard. This almost miraculous world of digital technologies continues its relentless march in broadening and deepening both methods and sources in the ambit of scholarly work. Scholars of the future will have access to amounts and kinds of information that we can only dream about today. Nevertheless, we need to think strategically about developing our IT capabilities and moving ATMA into the digital future. Without a clear vision and a rapid deployment of proactive strategies to create databases, ATMA will be pushed more and more to the side or even to the bottom of the global information infrastructure.

So far, we are proud of what we have achieved. Our databases are diversified, and our work is becoming better known among the right public. Updating the present collections, incorporating new technological realities and expanding our focus from the first database, *PADAT*, to other related collections, is a Herculean task. We have succeeded in persuading UKM's teaching staff to have their publications about the Malay World published in *PADAT*. In order to expand and diversify the databases, we have lifted our sights towards institutional and individual libraries outside UKM. Maintaining the perishable goods that we have created makes it necessary for us to assume a leadership role in managing information and in forming new alliances all the time all too conscious that user expectations may exceed our financial and technological ability to deliver. More importantly, we must put the interests of our customers first, since they are the ones who will be paying for maintaining and improving the products. New search strategies must be devised to enable busy researchers to obtain the right information at the right moment. Already, the public views ATMA in a light different from before, at least before April 1999. We have provided a blueprint for future development, all the while attempting to see far ahead, to show the way, to identify trends and to understand the implications. In this way, we seek to give ourselves momentum in seizing opportunities, and recognising and avoiding potential problems in time. Our achievements thus far are the result of an efficient working team, a visionary leadership, and overwhelming support from UKM's top management. We hope to develop our databases to meet world standards, and in doing so, make ATMA the defining point in the study of the Malay World.

ACKNOWLEDGEMENT

This is the original paper presented at “The 3rd UKM-UC International Conference on Societies for a Sustainable Future”, on 14-15th April 2003, at University of Canberra, Canberra, Australia.

REFERENCES

- Campbell, Jerry. 1995. Copyright, Public Policy and Digital Libraries: Searching for First Principles. IN Michael Mathews and Patricia Brenann, edited. *Copyright, Public Policy and Scholarly Communities*. Washington, D. C.: Association of Research Libraries.
- Detrie, Jean-Pierre & Ramanantosa, Bernard. 1986. Diversification – the Key Factors to Success. *Long Range Planning* 19(1): 31-37.
- Ding Choo Ming. 2000a. Access to Digital Information: Some Breakthroughs and Obstacles. *Journal of Librarianship and Information Science* 32(1): 26-32.
- _____ & Supyan Hussin 2000b. Preservation of Culture and Knowledge Through Technology: the Experience of UKM. A paper presented at *UKM-UC Seminar on Technology in Development*, 25th January, at UKM, Bangi.
- _____. 2002a. Indexing In Malay World Studies Data-Base: Some Critical Roles of Resource Centers in the Context of Knowledge Work and knowledge worker. A paper presented at *Seminar at UNIMARA*, Shah Alam, September.
- _____. 2002b. Access to Materials in and on Malay Studies from Leiden to Bangi: a Model of Information Repackaging on Information Superhighway. A paper presented at *ATMA-KITLV Colloquium: Dutch Scholarship and the Malay World: a critical assessment*, at UKM, 20-22 November.
- _____. 2002c. Accessing the Malay World. *SEAREP Bulletin* 7(1): 19-21.
- _____. 2002d. The Pantun Database of the Malaysian Baba. *SEAREP Bulletin* 7(2): 24-31.
- _____ & Arba'eyah Abdul Rahman. 2002e. Daripada Kamus Peribahasa Melayu Digital ATMA kepada Perkamusan Bahasa Berkomputer. *Dewan Bahasa* 2(7): 48-59.
- _____ & Supyan Hussin & Choo Wou Onn. 2002f. Pembinaan Pangkalan Data di ATMA; Tapak Kecil dalam Langkah Besar Membaiki Aliran Maklumat Sehala dari Barat ke Timur. A paper presented at *Seminar Kebangsaan Bahasa dan Pemikiran Melayu, Akademi Pengajian Melayu*, Universiti Malaya, 18-19 June.
- _____. 2003. The Malay World Portal www.malaycivilization.com.” *SENADA: Berita Universiti Kebangsaan Malaysia* 252: 10-11.
- Healy, S. A. 1998. Science, Technology and Future Sustainability. *Futures* 30(1): 23-42.
- Mathews, Linda M. 1998. The Collection Management Perspective. *Collection Management* 22(3-4): 193-197.
- McMullen, B. E., Masullo. M.J., IBM Corporation, & Postlethwaite, B. 1999. Digital Libraries: Potential and Risks. <http://www.educause.edu/ir/library/html/edu9948/edu9948.html>. 27 November 2002.
- Shamsul Amri Baharuddin, Ding Choo Ming, Supyan Hussin, Afendi Hamat. 2001. A Retrieval System for the Database on Malay World Studies. A paper presented at

- the *National Conference on Research and Development at Higher Institutions of Learning*, 25-26 October, PWTC, Kuala Lumpur.
- _____, Rumaizah Mohamad & Haslindawati Hamzah. 2002. Pengajian Alam Melayu di Pentas Global: Teknologi Maklumat dan Penstrukturan Ilmu di ATMA, UKM. A paper presented at *Seminar Kebangsaan Bahasa dan Pemikiran Melayu*, Akademi Pengajian Melayu, Universiti Malaya, 18-19 June.
- Supyan Hussin, Ding Choo Ming, Afendi Hamat, Kamariah Abu Samah & Hazrul A. Halim Nasir. 2002a. Development of a Database on Malay Studies at ATMA-UKM to Enhance Information Sharing and Preservation. *IRPA Report No: 04-02-02-0040. ATMA-UKM.*
- _____, & Arbai'eyah Abdul Rahman. 2002b. Sistem Capaian untuk Kamus Peribahasa Melayu Digital Pertama. A paper presented at *Seminar Kebangsaan dan Pemikiran Melayu: Kecemerlangan Melayu dalam Era Teknologi Maklumat dan Komunikasi*, 18-19 June. Akademi Pengajian Melayu, Universiti Malaya.
- Turner, R. K. (editor). 1993. *Sustainable Environmental Economics and Management: Principles and Practice*. London: Belhaven Press.
- Wilson III, Earnest J. 1998. Inventing the Global Information Future. *Futures* 30(1): 23-42.
- Sector, Glen M. 1997. Fair Use in Pay-Per-Use World. *Library Acquisitions, Practice & Theory* 21(1): 53-59.

Supyan Hussin
Felo Penyelidik Kanan
Institut Alam dan Tamadun Melayu
Universiti Kebangsaan Malaysia
43600 UKM Bangi
Selangor Darul Ehsan
e-mail: supyan@pkrisc.cc.ukm.my

Ding Choo Ming
Felo Penyelidik Kanan
Institut Alam dan Tamadun Melayu
Universiti Kebangsaan Malaysia
43600 UKM Bangi
Selangor Darul Ehsan
e-mail: chooming@pkrisc.cc.ukm.my