Re-emergence of CDS/ISIS and Scope for Cultural Information Systems: Unicode Compliance and Current Relevance in Indian Context

Hussain K.H

Abstract

Paper traces the distribution history of CDS/ISIS in India in the nineties, its wide acceptance in the early period and its ill-fate in later years. In the last decade ISIS has re-emerged as an Integrated Library Management Systems (ILMS) with the launching of ABCD. Also J-ISIS has emerged as an important tool assimilating advanced language and web technologies to create textual information systems with the same ease and vigor of old CDS/ISIS. In this transformation it has become Unicode compliant and replaced database engine with Berkely DB. Inverted File indexing has been updated with Apache Lucene. Paper briefly describes ISIS Software Family, its main components and related technologies. Possibility of using vernacular scripts in textual information system can bring out opportunities for building up heritage information systems. This will result in the increasing role of ISIS software in cultural documentation. The present study reports the development of a bibliography in Malayalam using J-ISIS, for the first time in an Indian language. In this endeavor LaTeX , the renowned typesetting package is used. The symbiosis between J-ISIS and LaTeX can open new horizons in preparing bibliographies in regional languages of Afro-Asian countries, especially in projects like Indian National Bibliography.

Keywords: Micro-CDS/ISIS, J-ISIS, ABCD, Cultural documentation, Heritage Information Systems, LaTeX, Typesetting Bibliographies, Textual Information Systems, Unicode, language Technology

Introduction

‘Computerised Document Systems/ Integrated Set for Information Systems’, popularly known as CDS/ISIS was introduced in India in the mid-eighties. Librarians in India welcomed it wholeheartedly. Developed and maintained by UNESCO, National Information System for Science and Technology (NISSAT), New Delhi took responsibility to distribute it free of cost in the country. Many training programmes and workshops were conducted and eminent teachers like Prof. Neelameghan were the resource persons. The same fervor was seen in Pakistan, Bangladesh and Latin America.

Many reasons can be attributed to this enthusiasm. IBM-PC was appearing on desktops at the time of its distribution in the third world. Librarians were hearing stories from the west about the usage of computers in libraries. Library science
departments in universities were renamed by adding ‘Information’ to their names to indicate the increased role of computers in bibliographic activities. Machine Readable Catalog (MARC) bibliographic standards were introduced in the curriculum along with Anglo American Cataloguing Rule (AACR). Creation of catalogs and printing of index cards were started using computers and dot-matrix printers in western countries in a big way and were called ‘electronic catalogs’ at that time. Terms like ‘digital’, ‘information technology’, etc. were yet to come.

CDS/ISIS started its successful journey in Indian subcontinent at a time when a few other packages available were beyond the reach of librarians. Even well-known academic and research libraries were running on meager funds at that time which was a major reason for immediate acceptance of the freely distributed Micro CDS/ISIS from UNESCO that excited them.

CDS/ISIS was more than a tool for electronic catalog from the very beginning. As the first part of its name implies, it was for ‘Computerised Document Systems’. Librarians were captivated by its capabilities to build up beyond bibliographic systems. As the second part implies it was an ‘Integrated Set for Information Systems’. As a database management system without demanding any programming skills for its adoption and usage, CDS/ISIS opened up a new world to librarians. Besides, they could attempt non-bibliographic systems too.

Even after three decades of its existence, CDS/ISIS exhibits its splendor in creating brilliant textual information systems without writing a line of code. Since numerical data handling is beyond its scope, information technology (IT) students have not heard about it. They are mainly trained in numerical relational databases and they often fail to conceive textual databases. But librarians with experience in CDS/ISIS, excel in this field. A program like ‘Google’ is complicated and gigantic for an ordinary IT student, but in reality, it is only an extended library catalog that indexes ‘sites’ instead of books.

In the first decade after the introduction of CDS/ISIS, beginning from 1985, thousands of librarians in India adopted micro-CDS/ISIS for automating their libraries. Small as well as big libraries started to adopt the package which resulted in forming of many informal discussion groups among librarians. The librarians began to consider it as the last word in library automation. Many practitioners had later attempted programming with its Pascal interface which gave librarians a real feeling of IT expertise.

(Index creations using Apache Lucene engine has been known in IT field during the last decade, while librarians were handling inverted files from 1985 onwards!)

**Retreat of CDS/ISIS**

By the later half of 90’s CDS/ISIS began to retreat from the scene. Librarians had grown beyond electronic catalogs and inverted index. A total library management system was what they were craving for. National as well as international IT firms recognised the market scope for such systems. On the other hand, UNESCO did not exploit the situation to develop an integrated library management system based on CDS/ISIS. MS Windows based ISIS named WINISIS had been launched by this time but it was only a replica of MS DOS package. NISSAT attempted a library management system called ‘Sanjay’ without any impact (Anitha Malik, 2013). Librarians were under pressing needs for a software package that can manage library housekeeping activities also. Within a few years, proprietary packages like LibSys and Alice could effectively conquer the software base built by CDS/ISIS. Some of the university libraries started to install SOUL developed by University Grand Commission (UGC/INFLIBNET). Those who loved and advocated CDS/ISIS were helpless to defend its cause and application (Hopkinson, 2009);( Lakshmana Moorthy, 2004).

---

**Re-emergence of CDS/ISIS: Unicode Compliance and Current Relevance**
Locally made packages were also on the run (Packages like LibSoft, BookMagic, etc. could catch up hundreds of academic libraries in Kerala, a small southern state in India). Free and Open Integrated library management systems (ILMS) like KOHA had to wait till 2005 to spread though it was launched in 1999 (Chouhan, 2010). By that time CDS/ISIS had almost retreated from the Indian scene. Most of the databases were either discarded or partially exported to proprietary packages. Catalogues of English books and documents were lucky, thanks to its formatting capabilities to produce textual outputs for migration. Those catalogs made for local titles were useless because of non-standardised Roman transliteration of vernacular script practiced by librarians. A minority of librarians continued to keep the passion for ISIS to create printed bibliographies (Indian Forestry Abstract, Kerala Forest Research Institute) and online catalogs (Raja Muthiah Research Library, Chennai).

By the time the full-fledged ABCD (‘Automatización de Bibliotecas y Centros de Documentación’ in Spanish, which means: Library and Documentation Centers Automation) developed by BIREME, Brazil made its appearance in 2011 with all the beauty and strength of ISIS, librarians had already been distanced from ISIS packages. Only UNESCO should be held responsible for the loss of a mass base built up by a national agency and thousands of librarians for a decade across the country. Disappearing in the midway to make librarians hopeless and confused, it even paved the way for a propaganda that it was a design of the west to sabotage growing bibliographic management systems in India and other Afro-Asian countries.

Lack of concern from UNESCO after 1995, non-updating of WINISIS with newer versions of Windows, time-lapsed launching of ABCD, taking out its practical sessions from library science curriculum, spreading of KOHA …and many other factors supplemented to the retreat of CDS/ISIS.

**Rise of Local Languages**

At the same time a significant event happened in Indian IT field – Unicode embedding of local scripts in MS Windows in 2003. Indian languages have been waiting for this technological development for years. A lot of activities gained momentum in the area of language technologies for its advancement. At the user level blogging and e-mailing were shifted to local scripts within no time. Language computing that was limited in word-processing and desktop publishing (DTP) for 25 years could suddenly think of creating database management systems using local script. Millions of web sites in local scripts adopted Unicode fonts. Unicode enabled Indian languages in building up textual information systems. Unfortunately no attempts were made to make WINISIS Unicode compliant. An ASCII based localization called MISIS (Malayalam ISIS) was attempted by the author in 2003 to catalog Malayalam books in colleges in Kerala (Raman Nair, 2011). It could render and retrieve records using Malayalam script but Unicode embedding of Indian scripts outmoded this successful experiment.

**Future of Textual Information Systems**

ISIS have made a feeling among some IT experts that its philosophy and approach cannot be ignored as a mere librarian’s package. Primary reason for its growing significance is its simplicity in materializing an information system. It comprises all modules necessary for structuring, modifying, applying and maintaining such systems. Queries can be made for in-depth search and output can be made in whatever way one wants. All these are achieved by not even writing a line of code! IT experts have often been taken aback by its power of creation of a full-fledged system and its search
capabilities. Linearity is fundamental to its way of looking data and defining database. It doesn’t entertain any number of tables but an ‘integrated database’ with a single table. Since it is fundamentally bibliographic, its repeating fields and subfields are inherited from the age old cataloging practices. Yet it gives liberty to define non-bibliographic systems. When it comes to library practices it even gives freedom to disregard MARC!

The ISIS software family has now fully adopted UNICODE and multi-linguality, which best suits its special mandate to promote self-reliance and knowledge-creation-by-sharing with the librarians and information workers (BIREME 2010). De Smet (2008a) considered that future developments of ISIS will no longer be classified simply as Database Management Systems, but will become the de facto standards for the management of textual, semi-structured databases. This new generation will be enriched with XML as a storage and exchange format. They will all use Lucene indexing and will all support Unicode.

A tool for cultural activism

ISIS was ‘free’ from the very beginning and its recent developments have become truly ‘Free and Open’. It enables non-IT people to create information systems as excellent as systems made by experts using gigantic DBMS. You need not be even a librarian to practice it. Science and technology personnel are likely to be closer to developments in IT, but those from arts and literatures are often ignorant of advancements in content creation. Compared to the domains of science and technology, cultural fields and activities lack funding and initiatives to create information systems. In case of diversity, richness and self initiatives, cultural activities are far and above science and technology. They deal with history and aesthetics and concerns about the existence of human beings and nature. Heritage and nature conservation are its prime agenda while ‘development’ without caring them are the present motives of technology. Vast and innumerable information-bases should be built to strengthen cultural activism without which this planet and human race will not survive. Science and technology owe to nature, history and heritage for cultivating revolutionary ideas and inventive minds.

ISIS and its new developments respecting local scripts have got more potential and reach than other DBMS in fostering information systems. From the very beginning, it spread in the so called ‘developing countries’ where most cultures are born but least documented. Cultural activism needs thousands of textual information systems, all are diversified and rich in content. They exhibit peculiar knowledge spaces often not logical to an ordinary IT expert, but well systemized to an activist in the field. Such content management systems can only be rightly conceived by these non-IT-experts. A tool like ISIS, free but powerful to create Information systems are rare in IT field. Specific and peculiar local systems can be built using local script and can be integrated to specialized systems nationally and globally. All these can effectively be constructed without a huge relational DBMS. That is why, Oracle has identified the significance of Berkeley DB that eventually became the database engine for recent ISIS developments.

The ISIS software family has a unique concept and importance both in its technological (to cope with Information Storage and Retrieval Systems) and users-community features (particularly for developing countries where the technology is widely known and used). The users community fully recognizes the enormous merits of UNESCO in developing and supporting the software, but fears that in the near future this might not be sustained; such change in policy would be seen as an abdication of its mission certainly at a point in the history of the software where important new developments require renewed efforts in
promotion, training and securing sustainability (BIREME, 2010).

Evolution of ISIS Family

‘Integrated Set of Information Subroutines’ (ISIS) developed by International Labour Organization in 1960, was the earliest package of ISIS software family. In 1975 UNESCO refined it as Micro CDS/ISIS for preparing a database of UNESCO publications. Later a modified version known as MINISIS was developed by International Development Research (IDRC) for use with minicomputer HP 3000. In 1986 UNESCO issued an ISIS version for use on IBM PC compatible microcomputers.

ISIS software evolved during the last six decades through different implementations, under numerous operating systems, programming languages and functionalities. In addition to UNESCO different organizations like IDRC, Canada and BIRME, Brazil sponsored its research and development.

Branching in many ways and intertwined often, ISIS Family is now thriving with vigorous growth adapting to Unicode and web technologies. Its evolution during last four decades is summarized in Figure I

The most exciting of these developments are J-ISIS and ABCD. J-ISIS recreates the same feeling of that good old days of 1986 with advanced language technology and web interface. Those who lamented on the extinction of CDS/ISIS from their life have got one more chance to sit with it and create wonderful information systems with the same wonderful formatting language. Those who were disappointed of not having a full library automation package based on CDS/ISIS do own now a great ILMS – the free and open ABCD. Though it has come at a time when all major libraries are succumbed to KOHA, ABCD excels in many aspects, mainly in the freedom of choice of basic definition other than MARC.

Egbert de Smet brilliantly expands these ideas (De Smet 2010). “Apart from its image as ‘computer science by and for information workers’, ISIS has another telling characteristic which also originates from its UN basis, i.e., the image of the typical...”
developing world user environment. Just like in car-production Volkswagen was said to be ‘Mr. Nobody’s Porsche’ (but we know how this story ended recently). ISIS was sometimes seen as a ‘poor man’s library system’. ISIS, in fact, is not even a library system, but a general-purpose tool allowing poor librarians to develop their own sustainable databases to semi-automate their libraries. In academic publications ISIS is often compared unfavorably with KOHA — a comparison which now needs to be made again with the arrival of ABCD! — and was seen as a poor substitute for the ‘real thing’. Even from a simply technological point of view we think this judgment is unfair. Social reality and image of ISIS as a tool for librarians will be very difficult to eradicate.”

He has summarized the advantages of new ISIS-technology as follows (de Smet 2009):

- It will have a flexible architecture in which ‘ISIS-cells’ will communicate through known protocols with several platforms and interfaces;
- The new ISIS generation will avoid the limitations related to database-, record- and field-sizes;
- ISIS databases will be Unicode compatible; and
- It will use Lucene from Apache Software Foundation for full-text indexing (i.e. The creation of inverted file generation).

J-ISIS: The Client-Server Java Implementation

An older-generation ISIS for Java (JavaISIS), made available in 1998 by DBA in Italy, should not be mistaken for the new J-ISIS which is entirely a new ISIS for Java. It is a new multi-platform FOSS ISIS software package which solves the limitations of its previous generation. J-ISIS is characterized by its sophisticated and flexible client-server application, implementation of Unicode, and its usage of a NoSQL database (Berkeley DB) and a new technology named Lucene for indexing and searching (Hopkinson, 2008).

Technologies behind J-ISIS

NoSQL Database: According to Oracle (2011), unlike RDBMS the process of supporting hundreds of online users for updating and reading from a very large dataset is accomplished by NoSQL databases by horizontally scaling across large number of servers. This will distribute the application load into several servers. The use of NoSQL database technology flattens the non-linear increase in total system cost curves and asymptotic degradation of performance with RDBMS technology (Couchbase, 2011).

Berkeley DB: J-ISIS uses Berkeley DB to overcome the size limitation of records. Berkeley DB is an Open Source embedded database system with a number of key advantages over comparable systems. It is simple to use, supports concurrent access by multiple users, and provides industrial-strength transaction support including surviving system and disk crashes. It can be used in applications that require high-performance concurrent storage and retrieval of key/value pairs. The software was originally distributed by Sleepycat as an Open Source product and provides a variety of programmatic interfaces (Olson et al., 1999). Now it is property of Oracle but keeps its free and open source nature.

Lucene Apache: Lucene is a Java based search engine library, which is an open source and freely available and can be integrated into different applications using Java method calls (Prasad & Patel, 2005). Indexing and searching are the two main functions supported by Lucene and J-ISIS uses these facilities. Lucene passes through different steps to index by encapsulating the original document, tokenizing process, parsing stop words, recording the frequency counts, sorting the postings, encoding the term positions, and storing them onto the disk (Carpenter, 2011)
Web-JISIS: It provides access to the databases in JISIS transferring a client request via TCP/IP protocol. It enables the end-user to use a web-browser to open a J-ISIS database, search resources, browse resources and to edit resources. Web-JISIS enhances the flexibility of J-ISIS benefiting from the advantages of web technologies. (Dauphin, 2010):

**ISIS and J-ISIS: Similarities and Differences**

Fundamentally, J-ISIS is the successor of CDS/ISIS for Windows in many respects. It is a general-purpose interface to manage ISIS-databases, and fully graphical. It implements all CDS/ISIS formatting language features and other functions like import and export. It presents itself as a single interface without a need for a web-browser or web-server. The fact that J-ISIS maintains ISO 2709 as a common data exchange format and enriched with XML as a storage and exchange format enables librarians themselves to use this technology without the assistance of IT-experts. However, there are significant differences. J-ISIS is Java-based and Unicode compliant and can run on all computer platforms without a need to convert the databases to the proper platform as is the case with classic ISIS. It is client-server, which allows access to remote databases through internet in addition to local databases. Instead of the classic MST-XRF, it uses Berkeley DB technology which will overcome the limits on current record and database-size and hence allow digital libraries to keep long full-text documents. Lucene indexing instead of the classic ISIS Inverted File, allows ranked search results on top of Boolean and full-text searching, but based on the classic Field selection (FST) of ISIS.

Above all existing applications will be compatible with new technologies with minimal effort of export-import. ISIS is no longer a ‘database’ but has become a standard for handling semi-structured textual databases (Hopkinson, 2008; ISIS3WC, 2010).

**Rio Declaration on the Future of ISIS**

The third World Meeting of ISIS (ISIS3WC) took place in Rio de Janeiro from 14 to 16 September 2008, organized by BIREME/PAHO/WHO with more than 150 participants from 31 countries and 4 continents representing tens of thousands of ISIS users worldwide. It addressed the current state of the ISIS Software Family as well as new important developments towards updating its platform. It affirmed that the ISIS Software Family has a unique technological concept and developmental mission to cope with information storage and retrieval systems particularly for developing countries where the technology is widely known and used. The conference affirmed the continued value of the facilities offered by the ISIS Software Family and decided to establish an international steering committee accredited by UNESCO. It also appealed UNESCO strongly to pursue its commitment that faded from its agenda for a while which affected its growth, distribution and acceptance.

Helen Hagos Berhe underlines these facts in her thesis (Berhe, 2012):

“Following different tests I conducted it has been proven that J-ISIS indeed can do more than traditional library applications. It provides high storage capability, Unicode compatibility, full-text indexing, and can handle a diversity of metadata and document formats. Study has identified some additional features which are very important to make J-ISIS fully capable software for open archives and metadata harvesting (OAI-PMH)… More investment and attention towards J-ISIS is quite important to benefit fully from the software in view of the already available features and capabilities. . . . . As a classic ISIS-technology, the ABCD library
software was enabled for digital library functions unlike its traditional library automation of bibliographic data.”

**Indian Scenario: A Case Study**

**Bibliography of Malayalam Books** *(Malayala Grandha Soochi - MGS)* of Kerala Sahitya Akademi is a unique bibliography in Indian languages which is a compilation of records of published works since 1772 in Malayalam, the language of Kerala, southernmost state in India. Compiled by K.M. Govi, former Head of Malayalam section of Indian National Bibliography, Calcutta, the seven printed volumes containing bibliographic details of 50,000 books were brought out by the Akademi. Only after 18 years the 8th volume for about 6000 books published during 1996-2000 could be compiled and printed. Unfortunately no attempts were made to assimilate modern language technologies and DBMS practices in the making of this volume. The same manual mode with index cards that K.M. Govi practiced were employed in 2016 also. The only touch of IT was with Adobe PageMaker for the final layout for printing. 6000 records in ASCII that spreads around thousand printed pages were in no way capable for either converting to a database or hosting in web.

The present MGS Committee of the Akademi guided by Prof. Lalitha Lenin (Retired Professor, Library and Information Science Department, Kerala University) who took in charge to compile the 9th volume (of period 2001-2005 ) started to reorganize the work with a different but clear perspective. She instructed that it should not be a compilation for mere printing as in previous years, but should be made in to a web-hostable bibliographic information system at the end of typesetting, fully functional with Malayalam script. Author was given full freedom to investigate appropriate technologies to achieve this.

After many years of personal detachment with CDS/ISIS author started to collect information from the net. The current status it attained during the last decade and the current state of developmental activities going on around was astounding. It is learnt that ISIS has assimilated all the advancements in Unicode and web technologies and matured to create bibliographies using any scripts in the world with the same good old vigor with Roman script. It is investigated that Unicode language technology attained in Malayalam since 2004 is fully functional with J-ISIS and MGS has got no other better option than J-ISIS. Data entry is progressing for nearly 7000 records and once the final output for typesetting the bibliography is produced by multilevel sorting and formatting, there exists Web-JISIS ready for hosting the content in net!

**New Indian Possibilities**

ISIS grown to Unicode and web cannot be simply limited to bibliographies. KOHA is widely installed in India but significance of ABCD as ILMS as an alternative for creating non-MARC systems. Moreover, ABCD is more amenable to configure to village and school libraries. (Kerala, southernmost state in India, alone have 5,000 village libraries and state government took initiative to automate them using KOHA, but not materialized in 10% of them after five years).

Significance of J-ISIS lies beyond Bibliographic systems. Similar packages to structure heritage systems with textual contents satisfying all its multitudes and specificities are rare. Since all the 22 official languages of India has attained Unicode technology the possibilities of making information systems in each language and state are hundreds of thousands. Traditional systems in Medicine, Astronomy, Mathematics, Folk Arts, music, performing arts, Architecture, . . . it is beyond bounds and limits. Majority of information systems in these fields are going to be created by peoples on initiatives. Big DBMS and heavy programming are out of question for these projects since funding from government or
other agencies cannot be expected. There lies the scope of J-ISIS to accomplish such democratic participations similar to the history of Wikipedia. MediaWiki is excellent in creating and relating primary documents in a subject field while J-ISIS will be adjudged to create information systems based on structured data filtered from them. MediaWiki creates HTML and XML documents and is, therefore, not expected for Boolean search and retrieval which is the realm of ISIS. Data can be taken out of records in any coded form, thanks to the formatting language of ISIS. This can be utilised to typeset bibliographies by exporting coded text to LaTeX which is considered to be the last word in scientific typesetting. It is interesting to recall author’s similar experiments between CDS/ISIS and the DTP package Xerox Ventura in the early nineties (Sarojam et al, 1995). LaTeX coding using ISIS formatting language is going to be implemented in an Indian languages with the current volume of MGS with the help of LaTeX experts. Symbiosis between J-ISIS and LaTeX will open new vistas to remold age old practices in preparing and printing bibliographies in Indian languages. The most benefitting agency from this methodology will be Indian National Bibliography. Once the bibliography is part of the LaTeX system, it can easily be ported to e-Pub, MediaWiki and XML which will have the same rendering and data structure similar to PDF output or can directly be derived from the XML backend of the ISIS itself. Bibliographies of any subject in any language can never dream such a multitude of existence beyond these.

Conclusion

Resurgence of CDS/ISIS from near oblivion is an exceptional case in the history of IT. From the very beginning its technical excellence and democratic role were identified by a few committed developers like Gianpaulo Del Bigio, Alan Hopkinson, Jean-Claude Dauphin, Egbert de Smet and Ernesto Spinak. When UNESCO began to retreat to the backstage in developmental activities J-ISIS and ABCD came to the forefront with more collaborative efforts and became Free and Open (FOSS). It is now poised for a cultural intervention in Asia, Africa and Latin America in building up systems using vernacular scripts. Descendant from a rich heritage, J-ISIS has eventually become in conserving heritage. It is a right thing to happen in IT and justifies its origin and evolution.

Acknowledgments

Prof. Lalitha Lenin gave me a chance and full freedom to choose and utilise J-ISIS for the creation of new volume of bibliography of Malayalam Books (MGS), Kerala Sahitya Akademy. C.V. Radhakrishan (Sayahna Foundation, http://www.sayahna.org/ and http://stmdocs.in/) who is an internationally known expert in LaTeX has agreed to assist in this process. Discussions with him exposed immense possibilities that arise from the symbiosis between J-ISIS and LaTeX.

Special thanks to Helen Hagos Berhe for her excellent thesis “Extension with Digital Library Technology of UNESCO’s J-ISIS Database Software” (2012) which documents the exciting story of revival of CDS/ISIS from oblivion. Her insight on the cultural importance of ISIS and plea to UNESCO to preserve and support ISIS inspired this paper. Production of a bibliography in my mother tongue Malayalam would not have been possible if I missed her thesis and knew ISIS is still there in the world for such tasks. It was finding a lost treasure after many years.

References


Re-emergence of CDS/ISIS: Unicode Compliance and Current Relevance

Universiteit Brussel. [Link](http://studyleib.net/doc/18155734/extension-with-digital-library-technology-of-unesco-s)

BIREME: The ISIS3 Conference (2010): New Challenges for a new future of ISIS


Hopkinson, Alan (2009): Library Automation in Developing Countries: the last 25 years. Information is for Development 25(4):304-312


