Development of relational construction for the GESR database

Yaser Mohammad Al Sawy *

Education Technology Department, Faculty of Education and Arts, Northern Border University, Saudi Arabia

A B S T R A C T

This study seeks to develop the infrastructure of the Gulf Electronic Scientific Research database (GESR), which depends on one of the integrated library systems that operate under the Machine readable cataloging (Henceforth, MARC21), The development of the database aims to optimize and expand the multi-language data display. As well as it enlarges databases reading resources of research systems in the Bibliographic Record MARC21, whether for basic search, advanced search, or specialized research. It also develops the capabilities of the fixed-length data field (008) to enhance the advanced search capabilities. The development system includes the expansion of crossing references to link the different forms of the main and added entries. The study aims at improving the application of RDA rules and the data display system according to the possibilities of the displaying form of the side sorting of the contents on the database website so that the lists of authors, title, keywords and biographies, Patents, institutions, and meetings. The research is based on the experimental analytical method, which depends on data analysis and processing through the integrated automated system, It presents the knowledge database as a unique system that combines the requirements of collecting and storing data availability and the technical work rules required by the MARC21 format. The researcher applied the analytical descriptive and the experimental approaches in order to identify the general and specific frameworks for the structure of the GESR database and to review the mechanisms for evaluating and developing the components of the integrated library system VTLS- VIRTUA.

1. Introduction

Databases are a good solution to the problems of the file system environment, most notably the problem of data redundancy and its negative impact on the use of storage media, conflicting information and the subsequent costs for its maintenance and operation. Moreover, updating any data item does not mean updating it at the system level, but it is limited to the file concerned with the update (Chavez-Gibson, 2013). This can result in incomplete data and the inability to impose centralized management and security control that protects information against interference or breach. The database can be defined as an electronic system designed to systematically collect data records and programs with providing minimal possible repeatability, more complementarity and sharing

data with different users without a link to data and the application programs.

The GESR Database seeks to create the first and largest database including Abstract and Full Text for the various forms of information resources from scientific research, academic thesis, articles, expert data files, patents and equipment files in the GCC countries in English and Arabic from Kuwait Institute for Scientific Research in state of Kuwait (PAAET, 2018), Databases are designed and managed based on the fixed scientific pillars of the creation and specialized competencies, which undergone multiple stages to reach the desired goal. All data are processed according to the MARC21 (Khurshid, 2002). The cataloging rules RDA (Resource Description and Access) are also adopted, which have been in place since April 2013, using the VTLS-VIRTUA system. It is one of the most important integrated library automated systems for building the database and displaying it on the Web Database Portal (Ghosh and Panda, 2011).

In this respect, the GESR database project aims at facilitating communication between the GCC countries' academic affairs and linking databases in universities, institutions and research centres

(Stilling et al., 2018). This will help to make a high progress in databases, higher education institutions and research centres in the GCC.

Based on PAAET (2018) the GESR database' goals are as follows:

1. Establishing a comprehensive database for the academic publication of researchers in the GCC countries in Arabic and English.
2. Providing data through a form of integrated automated library management systems.
3. The possibility of expansion of data size and the ability to apply the cataloging rules (RDA).
4. Technical flexibility to modify MARC21.
5. Design a comprehensive web interface to exhibit the database information.
6. The ease of use and comprehensiveness of research systems related to the database (basic - advanced - specialized).
7. The ability of the automated integrated library system to technically develop of the search in inactive descriptive bibliographic fields.
8. Flexibility in applying the personal attributes associated with the authority files.
9. Flexibility in developing the Crossing System for Authority Files.
10. Open access to data without any cost, barriers or blocking.
11. Comprehensiveness in the ability to search various data elements.
12. Transfer of technical expertise of the database internal development for all integrated library systems in MARC21.
13. The possibility of presenting the abstracts and contents in the search results.
14. Linking the research results to other bibliographic records (subjects - forms - date - lists of the involved authors - editions - series)
15. Developing the Hyperlinking feature with attachments such as CVs - educational films - web sites.
16. Minimizing the technical and financial effort to operate the GESR database.
17. Publication of the GCC countries' scientific research under a unified system and one browser.
18. The easiness and comprehensiveness of the database's technical development under a unified automated system and common technical services.

The components of the GESR database are as follows:

- Institutions data file: They are bibliographical records of the data of universities and institutions of higher education - and national research centres in the GCC participating countries.
- Experts data file: they consist of bibliographic records of researchers and academics' data and complete copies of their CVs.

- Information resources file: It includes researches, articles, books (reviewed and accredited) periodicals, facts, conferences papers and university thesis.
- Scientific devices file: it encompasses data and specifications of scientific and laboratory devices used in research.
- Patent file: it represents data, specifications and abstracts from the accredited inventions and inventors' data in the GCC.

The information resources database has several forms:

1. University Theses: they include master's and doctorate dissertations from universities or academic institutions.
2. Conference papers - Full or partial work of conference papers adopted at all conferences, seminars for researchers representing the GCC research institutions.
3. Scientific periodicals and journal articles.
4. Books - (that are reviewed, published and approved by the participating institutions in the database)
5. Reports and studies (issued by individuals and approved by the participating institutions in the database. Accredited scientific articles published in scientific journals of all researchers representing the scientific and academic institutions in the GCC countries.
6. Patents - (approved by official scientific bodies in the participating countries in the database).

2. Literature review

Han (2012), in his study entitled "Discovery New Control Bibliographic Library and Services" asserted that in order to meet growing needs of beneficiaries, access and retrieval of available resources on the web should be developed and optimized. In addition to the diversity of forms of catalogues and unified research applications, the used mechanisms by information centres should develop new search and retrieval tools. The study referred to the practice of modern applications specialized in access to sources and retrieval through the development of bibliographic records available in the information system or through the Metadata. Over and above that, the study displayed how to use these tools and the challenges faced by information centres in bibliographic settings, ensuring the quality of cataloging, levels of description and compliance with the bibliographic descriptive standards.

Jason (2011), in his research entitled **"Web Scale Discovery What and Why?"** pointed to the availability of the ability of cluster research systems to support the library environment by facilitating the linking of researchers to central digital warehouses, including Local content hosting. This allows accessing additional value-added resources such as searchlights, abstracts and available content in digital repositories. This present research focuses on

the introduction of cluster research systems and highlights key concepts of these systems.

The analysis of previous studies shows they bore some similarities with this study in dealing with the idea of developing research systems or analysis of the databases structure (Fitzgibbons and Meert, 2010). However, the present research is more accurate in addressing the real needs of beneficiaries to develop research systems and technical processing of the RDA (Park and Tosaka, 2015). In fact, it is an empirical study combining the analytical theoretical design with the comprehensive design and implementation of the database, including treatment through the MARC21 environment (Maier, 2011), in both languages Arabic and English.

3. Methodology

The researcher used the analytical descriptive and experimental research methodology to identify the general and specific frameworks for the structure of the GESR database and present the evaluation and development mechanisms. In addition to the implementation of the experimental study, the automated integrated library system's technical components VTLS –VIRTUA are improved. The technical authorization of the necessary modifications to the development of the database structure is carried out in full coordination between the technical team of the database with the system support team in the United States and its regional headquarters in Spain (Wang, 2009).

4. Discussion

4.1. The Virginia tech library system VTLS (VIRTUA)

The system was primarily developed by the Virginia Polytechnic Institute and the University of Virginia, USA. The system is named The Virginia Tech Library System (VTLS), it includes many subsystems. The Micro VTLS System is an integrated library system that is principally designed for small libraries and has great flexibility, while maintaining the integrity of the techniques and search capabilities of large VTLS systems (Rodgers and Puterbaugh, 2017).

VTLS Inc. is a private, profit-oriented organization based in Blacksburg, Virginia. It is a leader in the field of marketing, development and support of automated library systems as well as resource-sharing through computer networks and digital libraries. It has also obtained the ISO 9001 certification for its role in supporting and using the Z39.50 standard. It is also a member of several international organizations specialized in libraries Such as the International Federation of Libraries and Information (IFLA) and the American Library and Information Association (ALA). The VLTS system consists of eleven sub-systems, which are as follows: (Easy Cataloging System (Easy CAT) – Acquisition Subsystem– subsystem for series control- Inventory Subsystem - Online System on the Direct Easy PAC - Circulation subsystem - Booking subsystem – Study booking subsystem - Periodical indexing subsystem - Document delivery subsystem - Financial accounting system) (Tuna et al., 2017). VTLS operates under UNIX, IBM's VM operating systems and C, COBOL and All Base programming languages. UNIX operating system is installed on a Host Computer to which a set of personal computers are linked. They function by the Windows operating system where the search, input and technical services operations are conducted. Added to that, VLTS supports the Arabic language, which provides the following possibilities (VTLS ILS, 2018):

1. All screens can appear in Arabic including the lists.
2. Arabic can be used with English at the same time.
3. All messages and system updates of the system appear in Arabic while appearing in English when the language is selected.
4. Cataloging and bibliographic searches can be done in Arabic, English or both.
5. Searching by keywords in Arabic.
6. Availability of the Online Public Access Catalogue (OPAC).
7. Cataloging data management, which encompasses data input and modification, and MARC21 records.
8. Control the circulation of information resources, lending renewal, booking of materials as well as managing beneficiaries' records, automated lending operations and fines for materials delay.
9. The ability to write reports and prepare statistics. The system collects statistics on the circulation of information resources, fines for delays and violations of beneficiaries, details about the office groups, and statistics on the few titles that heavily circulate among the titles of the group.
10. Display the data via the LAN.

VIRTUA is the third generation of VTLS; it consists of object-oriented design and other technologies, including: Unicode support, Three-tier client server architecture, Rapid application development tools, Relational Database management systems, Stateless OPAC, and Unicode support (Bailey, 2011).

4.2. Developing the research systems of the GESR database

The search engine in the database is a technology that helps to access the information resources and consists of three main elements of the robot, Index and finally the Interface Search (Alyami and Assiri, 2018). The search engine develops methods of sorting results according to the research topic. Research strategy is one of the most important priorities for building integrated systems for library management (Scott and O'Sullivan, 2005) and database in order to ensure access to all information

that allow the researcher to have universal access to all elements of data (Kumar, 2012).

4.2.1. Basic search system development (Keywords)

A basic search is done by a researcher using a search term, whether it is part of a book, any form of headings, any form of topics, an author name, or a conference name. In fact, the search comprises any order in all parts of the bibliographic record entered into the GESR database (Kolowich, 2009). On this level, the development provides reading from all entered data into the database, where all titles (titles, authors, subjects) can be expanded into the MARC TAGS (Megnigbeto, 2012). Search fields in the title are also expanded, where keyword search can provide any part of those titles according to the following form of MARC21 (Zhou, 2013).

The search fields in the subjects have been also expanded to include keyword search for any of the topics according to the following form of MARC21:

Titles: (210 Abbreviated Title - 222 Key Title - 240 Uniform Title - 242 Translation of Title by Cataloging Agency - 243 Collective Uniform Title - 245 Title Statement 246 Varying Form of Title - 440 Series Statement/Added Entry-Title- 490 Series Statement - 505 Formatted Contents Note - 520 Summary, etc. - 800 Series Added Entry - Personal Name - 810 Series Added Entry - Corporate Name - 811 Series Added Entry - Meeting Name - 830 Series Added Entry - Uniform Title).

Authors: (XX0 Personal Names-General Information (100, 700) - X10 Corporate Names-General Information (110, 710) - X11 Meeting Names-General Information (111, 711) - 245\c - Statement of responsibility, etc. - 250\c - Statement of responsibility, etc. - 378\q - Fuller form of personal name - 490\c - Statement of responsibility, etc.- 505 Formatted Contents Note - 520 Summary, etc.).

Subjects: (600 Subject Added Entry - Personal Name - 610 Subject Added Entry - Corporate Name - 611 Subject Added Entry - Meeting Name - 630 Subject Added Entry - Uniform Title - 648 Subject Added Entry - Chronological Term - 650 Subject Added Entry - Topical Term - 651 Subject Added Entry - Geographic Name - 653 Index Term - Uncontrolled - 654 Subject Added Entry - Faceted Topical Terms - 655 Index Term - Genre/Form - 656 Index Term - Occupation - 657 Index Term – Function - 658 Index Term - Curriculum Objective - 662 Subject Added Entry - Hierarchical Place Name - 69X Local Subject Access Fields505 Formatted Contents Note- 520 Summary, etc.).

4.2.2. Development of advanced search system

The fixed-length data field is one of the most important of the MARC21 (Shen, 2016) where many of the fixed properties of the bibliographic record are encoded (Fig. 1), namely: cataloging level - date of publication - type of audience - form of governmental and non-governmental publications – the inclusion of bibliographic record to indexes and catalogues – language- place of Publication - Conference papers - Form of literary and artistic recordings - Format of applicable descriptive indexing rules - Notes type - Source of descriptive cataloging.

Fig. 1: The fixed-length data field

Based on the development of applications of all static field data input (008), the database is ready to provide researchers with accurate and specific results, which are sorted according to the previous sorting factors. This makes the development of advanced research comprehensive and complete (Dorta-González and Santana-Jiménez, 2017).

4.2.3. Development of the authority formats and the establishment of crossing

The development of fixed and standardized for the authority files format is an important technical tool in descriptive cataloging Technology provided the opportunity to quickly unify these formats while preparing all unused formats as non-authoritative files through the MARC21 authority fields (Library of Congress, 2018a).

Based on this, (Authors, conferences, bodies) are unified within the GESR database, where a single standardized format was adopted to link all the recordings or thematic treatments associated with the name of the authority file with the preparation of all other non-authoritative formats as search formats and the unified format link.

Therefore, these formats enable to search key words or lists to refer to the unified authority formats as shown in Fig. 2 and Fig. 3.
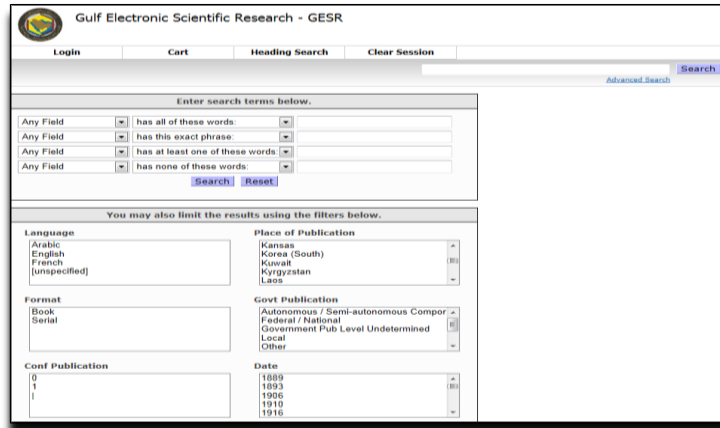

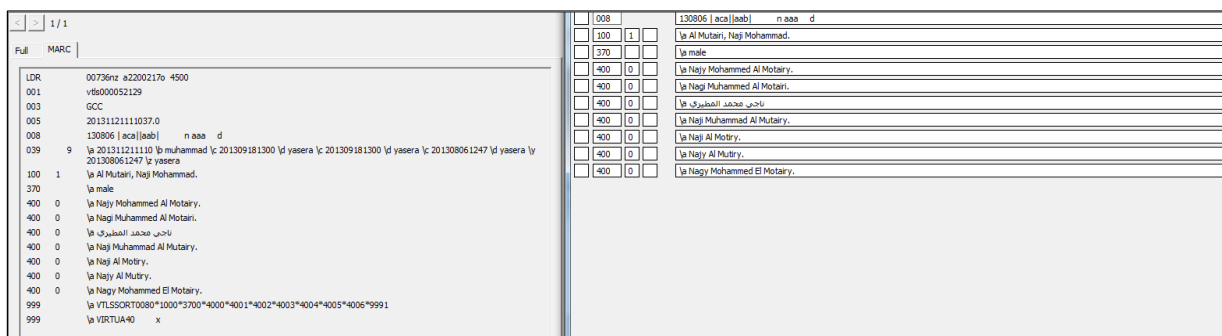**Fig. 2:** The GESR Database: The fixed-length data field

Fig. 3: The GESR database: Development of the authority formats and the establishment of crossing

4.2.4. Develop personal attributes file for the authority files

The creation of the Personal Attributes file within the GESR database is one of the new innovations under the RDA (Tosaka and Park, 2014), which has been in place since April 2013. Indeed, they enable to fully handle all data and personal attributes of individuals, entities and forums, according to the description fields mainly associated with the authority format, which allowed the development and construction of the expert file database (Library of Congress, 2018b): (370 (Associated place) - 371 (Address) - 372 (Field of Activity) - 373 (Associated Group) - 374 (Occupation)- 375 (Gender) - 376 (Family Information) - 377 (Associated Language) - 378 (Fuller Form of Personal Name) - 380 (Form of Work)- 381 (Other Distinguishing Characteristics of Work or Expression).

4.2.5. Development of the abstract format

The abstract is an important form of information that the researcher is keen to prepare them, because they lead to (Library of Congress, 2018c):

- Rapid access to information content
- The abstract includes many important scientific terms in the field of research
- The quick decision-making by the researcher in the use of research or not and, which saves more

effort and time to use the information source or search for another source.

The abstract often consists of 250 words that cover the following aspects: (introduction- problem of the study- purpose of the study- research questions – research hypotheses – the methodology used in the study and its tools – research results- research recommendations). The introduction of the abstract is a qualitative leap and a unique innovation in the GESR database, which enables the transition from bibliographic database format to abstracts databases in order to benefit more from the potential of the field 520 and its search ability. In addition, the abstract was developed within the GESR database in order to search through the keyword by means of any word or term it contains.

4.2.6. Development of the linkage system

The properties of the binding field 856 have been activated to display all the linkage points for expert files, online journal articles, and full text versions of forms of information resources (Fig. 4).

4.2.7. Develop the presentation of the search results

By clicking on any record to review its data, the data is arranged briefly for specific elements on the web browser including / (title - main portal - subject

headings - descriptors - website). Following the previous operation, the transfer to all other linked records under this entry (Figs. 5-8).

Fig. 4: The GESR database: development of the linkage system

Fig. 5: The GESR Database: The short search results

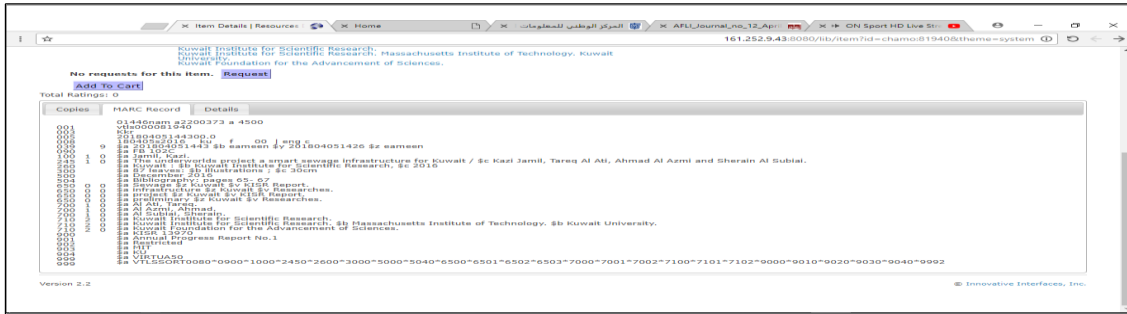**Fig. 6:** GESR Database: The complete search results

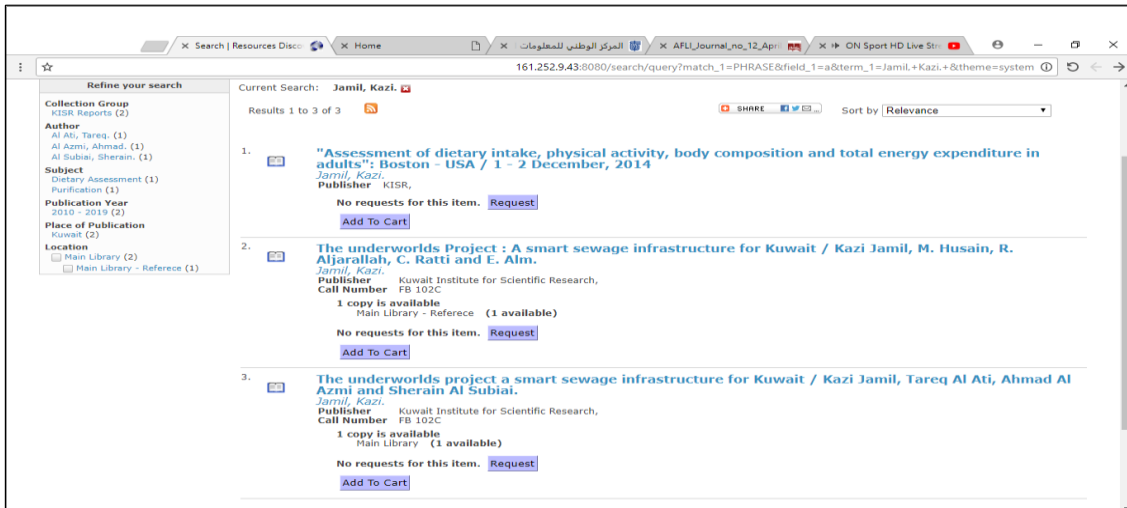**Fig. 7:** The GESR database: MARC record

Fig. 8: The GESR database: Sorting results

5. Research results

The present study led to the following results:

1. The GESR database is one of the most comprehensive research projects to inventory, store and retrieve data in a complete electronic format in the GCC countries.
2. The technical team of the database has perfectly chosen the VTLS, with its unique technical capabilities, which enabled the team to work effectively.
3. There is a real and integrated development of the database structure, combining the assimilation the technical work requirements with the beneficiaries' needs, while complying with the MARC21.
4. The technical team of the database has succeeded in developing the structure of (basic -advanced) search systems for reading from the unused cataloging fields of authors, titles and subjects.
5. The GESR database provided a complete processing of the fixed-length data field (008), which is an essential component of advanced search in sorting and identification.
6. The GESR database contains comprehensive and multiple components of information resource sections.
7. The database covers all forms of information resources.
8. The database is committed to the application of the RDA.

9. The database provides the technical support to the universities and research centres involved in the project, where the mind-set and work requirements are standardized.
10. The database provides for indexing of information resources in both Arabic and English.
11. The GESR database affords the processing of Personal Attributes.
12. The database presents a comprehensive system to link referrals between forms of authority files.
13. The database provides the linkage system for the attached files and the availability sites on the Web by expanding the application of field 856.
14. The GESR database provided a distinct availability of data presentation via the Database Portal. It included all of the data belonging to the non-specialist user and the comprehensive data presentation of the specialized researcher, as well as the MARC Format data display for the cataloging specialist that allows exporting bibliographic records.

6. Conclusion

The researcher conceptualizes that the success of the technical team in the development of GESR based on VTLS-VIRTUA means that all other systems (Unicorn – Horizon – Symphony – Sierra) operating through the MARC21 environment can achieve the same results by following the same technical treatment procedures. It is viewed that the success of the development is due to the advantages of the

VTLS system as well as the flexibility of the MARC21 cataloging fields and the advent in the RDA rules. Within each of these findings, it is identified that GESR database is due to the in the quality of the performance of technical procedures and comprehensiveness. However, the researcher postulates that there are shortcomings in the technical treatments of the comprehensiveness of cataloging in some fields such as thematic treatments, additional entries and the lack of linking references within the main entries for authors - bodies – forums. There is also a clear lack of Technical treatment for thematic link in 7XX fields, and linkage in field 856.

7. Recommendations

There is an urgent need for a comprehensive review of the bibliographical description comprising the added entries for (authors - subjects - institutions – meetings) Authority files of the subject headings and keywords have to be unified. Moreover, many aspects have to be completed such as the unified divisions of subject headings, the abstract for all bibliographic records and the thematic link reference in the 7XX fields. It is highly recommended to comprehensively treat and revise indicators in all cataloging fields and eliminate the field 260 like academic thesis. Over and above that, it became obvious to start Thesaurus (a list of hierarchical function keywords), as well as adjusting the thematic treatment of unifying divisions for spaces inside the subject headings as a basis for GIS linkage.

References

Alyami HY and Assiri EA (2018). Invisible web and academic research: A partnership for quality. International Education Studies, 11(4): 84-91.

Bailey K (2011). Online public access catalog: The Google maps of the library world. Computers in Libraries, 31(6); 30-34.

Chavez-Gibson S (2013). The comprehensive, powerful, academic database (CPAD): An evaluative study of a predictive tool designed for elementary school personnel in identifying at-risk students through progress, curriculum, and performance monitoring. The University of Texas at El Paso, El Paso, USA.

Dorta-González P and Santana-Jiménez Y (2017). Prevalence and citation advantage of gold open access in the subject areas of the Scopus database. Research Evaluation, 27(1): 1-15.

Fitzgibbons M and Meert D (2010). Are bibliographic management software search interfaces reliable?: A comparison between search results obtained using database interfaces and the EndNote online search function. The Journal of Academic Librarianship, 36(2): 144-150.

Ghosh TK and Panda KC (2011). Automated serials control at the Indian institutes of technology: An overview. Program, 45(2): 173-184.

Han MJ (2012). New discovery services and library bibliographic control. Library Trends, 61(1): 162-172.

Jason V (2011). Web scale discovery what and why?. Library Technology Reports, 47(1): 5-11.

Khurshid Z (2002). From MARC to MARC 21 and beyond: Some reflections on MARC and the Arabic language. Library Hi Tech, 20(3): 370-377.

Kolowich S (2009). In search of a better search engine. The Chronicle of Higher Education, 55(24). Available online at: https://eric.ed.gov/?id=EJ832340

Kumar S (2012). Impact of internet search engines on OPAC users: A study of Punjabi University, Patiala (India). Program, 46(1): 56-70.

Library of Congress (2018a). MARC 21 format for authority data. Available online at: https://www.loc.gov

Library of Congress (2018b). Fuller form of personal name attribute. Available online at: https://www.loc.gov

Library of Congress (2018c). Summary note. Available online at: https://www.loc.gov

Maier SB (2011). MPLP and the catalog record as a finding aid. Journal of Archival Organization, 9(1): 32-44.

Megnigbeto E (2012). A UNIMARC bibliographic format database for ABCD. Program, 46(4): 458-467.

PAAET (2018). GESR project. Available online at: http://www.paaet.edu.kw

Park JR and Tosaka Y (2015). RDA implementation and training issues across United States academic libraries: An in-depth e-mail interview study. Journal of Education for Library and Information Science, 56(3): 252-266.

Rodgers AR and Puterbaugh M (2017). Digital badges and library instructional programs: Academic library case study. Journal of Electronic Resources Librarianship, 29(4): 236-244.

Scott TJ and O'Sullivan MK (2005). Analyzing student search strategies: Making a case for integrating information literacy skills into the curriculum. Teacher Librarian, 33(1): 21-25.

Shen Y (2016). Strategic planning for a data-driven, shared-access research enterprise: Virginia tech research data assessment and landscape study. College and Research Libraries, 77(4): 500-519.

Stilling GES, Byrd AS, Mazza ER, and Bergman SM (2018). Still a deadly disease? Performance appraisal systems in academic libraries in the United States. College and Research Libraries, 79(3): 366-400.

Tosaka Y and Park JR (2014). RDA: Training and continuing education needs in academic libraries. Journal of Education for Library and Information Science, 55(1): 3-25.

Tuna T, Subhlok J, Barker L, Shah S, Johnson O, and Hovey C (2017). Indexed captioned searchable videos: A learning companion for STEM coursework. Journal of Science Education and Technology, 26(1): 82-99.

VTLS ILS (2018). VIRTUA library system. Available online at: https://www.iii.com

Wang Z (2009). Integrated library system (ILS) challenges and opportunities: A survey of US academic libraries with migration projects. The Journal of Academic Librarianship, 35(3): 207-220.

Zhou M (2013). A systematic understanding of successful web searches in information-based tasks. Journal of Educational Technology and Society, 16(1): 321-331.