

## The eBethArké Syriac digital library: a case study

Rutgers University has made this article freely available. Please share how this access benefits you.  
Your story matters. [\[https://rucore.libraries.rutgers.edu/rutgers-lib/51587/story/\]](https://rucore.libraries.rutgers.edu/rutgers-lib/51587/story/)

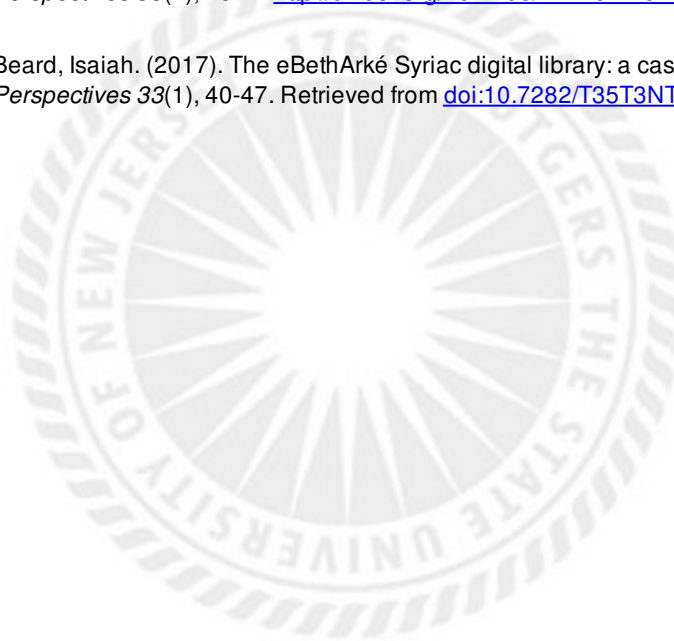
This work is an **ACCEPTED MANUSCRIPT (AM)**

This is the author's manuscript for a work that has been accepted for publication. Changes resulting from the publishing process, such as copyediting, final layout, and pagination, may not be reflected in this document. The publisher takes permanent responsibility for the work. Content and layout follow publisher's submission requirements.

Citation for this version and the definitive version are shown below.

**Citation to Publisher** Beard, Isaiah. (2017). The eBethArké Syriac digital library: a case study. *Digital Library Perspectives* 33(1), 40-47. <http://dx.doi.org/10.1108/DLP-07-2016-0017>.

**Citation to this Version:** Beard, Isaiah. (2017). The eBethArké Syriac digital library: a case study. *Digital Library Perspectives* 33(1), 40-47. Retrieved from [doi:10.7282/T35T3NTP](https://doi.org/10.7282/T35T3NTP).



**Terms of Use:** Copyright for scholarly resources published in RUcore is retained by the copyright holder. By virtue of its appearance in this open access medium, you are free to use this resource, with proper attribution, in educational and other non-commercial settings. Other uses, such as reproduction or republication, may require the permission of the copyright holder.

*Article begins on next page*

# The eBethArké Syriac Digital Library: A Case Study

Isaiah Beard

*Libraries, Rutgers, The State  
University of NJ, New Brunswick,  
New Jersey, USA*

## ABSTRACT

The eBetharké Syriac Digital Library Portal is a collaborative effort between the libraries at Rutgers, The State University of New Jersey, and the Beth Mardutho Syriac Institute, a traditional library of texts; to create a specialized digital library collection online. This digital library features content in and relating to Syriac, an Aramaic dialect spoken in the 1<sup>st</sup> century A.D. and for which a great deal of historically significant documents were written during the period. This task required effort and research on multiple fronts, including software development; collaboration on technical, interpersonal, and policy-based levels; and in overcoming challenges related to the predominant computing platforms installed and in use by potential users of this digital library. This collaboration provided significant new challenges and learning experiences among the staff who worked on this project and provides a base upon which our digital library platforms can diversify and be more culturally aware.

## General Terms

Management, Documentation, Performance, Design.

## Keywords

Digital Library, trusted platform, preservation, Syriac, Unicode, RUcore, Rutgers, Institutional repository, Beth Mardutho.

## INTRODUCTION

Syriac is a language, a dialect of Middle Aramaic spoken in the 1<sup>st</sup> Century A.D. over much of the Fertile Crescent, and becoming a major literary language throughout much of this same region from the 4<sup>th</sup> through the 8<sup>th</sup> Century A.D. ("Syriac," n.d.). Since then it has been mostly replaced by Arabic, and Syriac was a subject of minor research in Middle Eastern studies until the 1970s, but recently, literary scholars have taken a renewed interest in the language, and in literary works written in Syriac.

As interest has piqued, the need for more online resources for Syriac and Arabic texts has become a sore spot among scholars in this field. Syriac researchers find they are hampered in the age of digital scholarship by a lack of a supporting technological framework to conduct effective research in the ancient literatures written in this language, with most finding aids unavailable in online formats and authoritative Syriac Literature histories being over a century old (Gibson, et. al., 2016). Consequently, there have been increasing calls within the academic community for better support in the technological space for this and other contemporary languages of the region. Indeed, scholars of Middle Eastern texts would benefit greatly from a mass digitization project in much the same way Google Books and Hathi Trust have revolutionized online text access for the Western world (Heal, 2012).

A significant cachet of literary works written in and regarding the Syriac languages exists in Piscataway, NJ, at the Beth Mardutho Syriac Institute. At this location, a library of books and publications exist, many of which are out of copyright. The Institute has also been working diligently to bring native support for the Syriac language to most computer operating systems, making it possible to read Syriac texts online using Microsoft Windows and Linux without any additional software, and on current Mac OS X systems (with the installation of appropriate fonts). To develop this effort further, the group embarked on a partnership with Rutgers University Libraries beginning in the summer of 2012 to develop a digital library for their online works. An initial 1,500 objects have been identified and are being digitized for this purpose.

One major product of this effort has been the unveiling of the eBetharké Syriac Online Portal. A prototype is now online, representing the evolution of nearly a decade of collective web development in creating an online academic digital repository, and further adapting it to non-English languages.

The project's main purpose is to be a production-level, digital preservation-specific Syriac digital library, bringing a significant repository of works into the digital age and simultaneously providing a long-term digital preservation strategy. It supports the display of metadata and descriptive details for digital library objects not just in English, but in Arabic and Syriac languages as well, where appropriate. Extensive reliance on true Unicode support by the software is necessary and yet surprisingly found lacking by default in a number of software libraries required to enable the platform. Project team members had to rely on their own programming and debugging expertise to make the integration work seamlessly. Further significant testing of web browsers, and how our existing repository infrastructure accommodates

user-entered and XML-parsed Syriac search terms and data records, proved to be a significant hurdle for which there was limited history and knowledge to fall back upon.

## Background

Previous attempts had been made at creating an online repository for Syriac and Syriac-related texts based off the Beth Mardutho institute's existing collections. On a parallel track, separate efforts were underway at Rutgers, the State University of New Jersey, to develop a comprehensive digital library and institutional repository architecture. Ultimately, these paths converged, culminating in the partnership and digital library project outlined here.

### Early Attempts

The Institute's stated mission is to promote the study and preservation of the Syriac heritage and language, as well as to facilitate research opportunities in the field for the academic community and heirs of the Syriac heritage. It thus not only serves a library for works written in and regarding the language, but acts an advocate for Syriac in the academic and technological diasporas. Such efforts include not only scholarly research, but efforts to make the Syriac language readable on modern computer Operating Systems. To that end, the institute has developed OpenType fonts for displaying the language, and advocated for inclusion of the language in modern character sets. The Syriac language is now natively supported as a code page in the Unicode UTF-8 character set (Schild, n.d.).

The main collection at Beth Mardutho encompasses approximately 30,000 resources, primarily manuscripts printed or written in Syriac, Arabic, English, Latin, Aramaic, and Turkish. Efforts to make Syriac-related works available in digital form date as far back to June 2000, when the institute embarked upon a multi-institutional partnership for an electronic library portal (Kiraz, 2001). Partners included Brigham Young University, the Catholic University of America, Brown University, Dumbarton Oaks Library, and the Princeton Theological Seminar. This early attempt at a digital library ultimately yielded a basic online catalog which served the institute for several years. However, as time progressed, increasingly complex objectives and the need to grow the collection required a more extensible and adaptable platform.

### Rutgers University Libraries, and RUcore

The Rutgers University Community Repository (RUcore, located at <http://rucore.libraries.rutgers.edu>) is a preservation and access platform that stores and makes accessible digital research and educational materials created by the Rutgers academic community and its strategic collaborators. The goal of RUcore is to advance research and learning, to foster interdisciplinary collaboration, and to contribute to the development of new knowledge through the archiving, preservation, and presentation of digital resources. Original research products and papers of the faculty and administrators and the unique resources of the libraries are permanently preserved and made accessible.

RUcore emerged when the notion of digital preservation was at its infancy, required a great deal of real-world testing to be shown as trustworthy, and re-defined the process and steps necessary to seek to maintain items of historical value for decades. This feat requires overcoming major technical barriers and building trust within the community it serves (Jantz, 2005). At its core is an open-source digital asset management software platform, known as the Flexible Extensible Digital Object Repository Architecture (FEDORA). This software was developed by a team of researchers at Cornell University, and is a baseline platform for archiving, managing and making accessible a variety of digital assets, using an object/service-oriented framework based on architecture models developed by Kahn and Wilensky (1995). Building on this platform, RUcore utilizes object models, or preservation architectures which build upon layers of metadata and descriptive elements (Jantz, Giarlo, 2005). Each object contains a high-quality, stable preservation file that meets standards for longevity; a series of web-optimized presentation files; descriptive, rights and technical metadata; and when available, Optical Character Recognition output and other XML-based associated data.

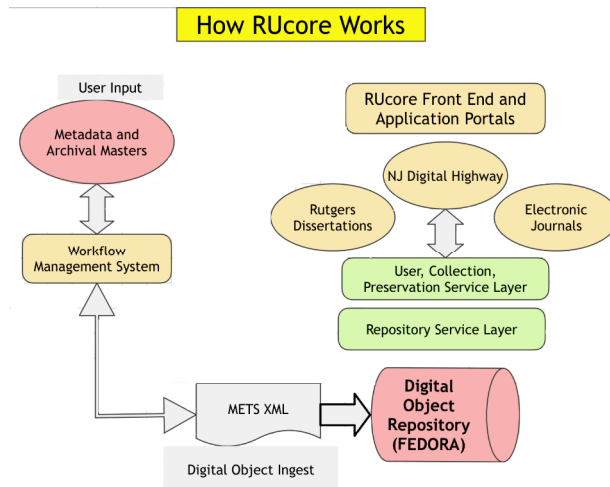


Figure 1: RUcore basic architecture.

Additional software was needed to provide a web-based method in which users could browse or search for objects stored in this repository. To answer this need, the Open Workflow Management System (OpenWMS) was developed. OpenWMS can be used standalone, or in conjunction with a FEDORA installation to provide many of the building blocks needed for a fully functional digital library. From the beginning, the development of this platform and associated software was based on a need to support management access and preservation of historic and cultural source materials (Agnew, Yu, 2007).

A corollary to RUcore's academic repository mission is a continued effort to reach out to nearby cultural heritage organizations, museums, libraries, and similar entities housing content of historic or cultural significance, and permitting use of the same technological platform underpinning RUcore's academic repository. A prime example of this initiative is the New Jersey Digital Highway (NJDH for short, accessible at <http://njdigitalhighway.org>). Founded in 2004, NJDH is an ongoing effort to reach out to identified cultural heritage institutions within the state of New Jersey, to catalog, and preserve their historic collection in a digital form (Jeng, 2008).

Whereas RUcore initially launched with a capability of preserving and making available still images, scanned documents and manuscripts, the platform evolved to support sound, moving images, and varied scientific research datasets as well. Notable examples of this expansion include the Video Mosaic Collaborative (VMC, accessible at <http://videomosaic.org>), an interactive collaboration interface designed to enable educators to utilize classroom videos to make discoveries in math education (Agnew, Mills, Maher, 2010).

After much due diligence, Beth Mardutho and Rutgers University Libraries signed a partnership agreement, pledging to help extend digital library functionality and technology to the Syriac collections. These digitized works will officially be hosted on RUcore, through a specialized portal presence that gives a distinct online identity to Beth Mardutho.

### *Existing feature set*

In placing Beth Mardutho's digital content on the RUcore portal, a core set of primary features will be immediately available to the project's stakeholders:

- **Preservation**  
RUcore is, first and foremost, a preservation platform for digital files and content. As our continual pledge is to evolve, migrate, and preserve all digital content stored on RUcore for as long as we exist as an entity, those same commitments transfer to the Beth Mardutho digital collections and content for as long as they choose remain a partner with the RUcore platform.
- **Digitization technology and facilities**  
RUcore's mission (and, by extension, Beth Mardutho) is further supported by facilities dedicated to the handling of digital objects, and the creation of digital surrogates of analog objects through digitization. To support this need, the Digital Curation Research Center (DCRC) is maintained in RUL's flagship facility on the Rutgers-New Brunswick campus. The DCRC maintains workstations, hardware and software devoted to the handling of digital resources destined to be preserved in RUcore. These include video and audio editing and transcoding; DVD and digital streaming file authoring; photograph, still image and book scanning; still image slide/film digitization and restoration; and the imaging and printing of large format objects (such as photos, posters and maps in sizes up to 3 feet by 4 feet, and with special handling, some objects of larger size).
- **Metadata and Cataloging/Database functionality**  
A requirement of the platform from inception is that it be flexible in the type of controlled vocabularies and text entries it accepts from catalogers and data entry staff. The designers settled on using MODS as the prevailing metadata standard for source and descriptive metadata. Controlled vocabularies such as the Library of Congress Subject Headings, Art & Architecture Thesaurus, and other common ontologies can be configured, and were used when relevant heading swere available for resources in this collection. Local and vernacular databases can also be loaded, if provided by the collection owner.

For this case, we consulted directly with George Kiraz, Ph.D., who runs the Beth Mardutho Syriac Institute as well as Gorgias Press, for subject headings and metadata vocabularies. His direct involvement in the earlier incarnations of this electronic library also afforded us access to his catalog and inventories, upon which a custom database and controlled vocabulary were devised to kickstart the library.

### *Required added features, and associated challenges*

By no means was adding Syriac and other Middle Eastern language texts to RUcore a trivial task. The majority of the development for this platform had been, up to now, focused on texts in English and other languages with a predominantly Latin character set. As such, most of the search engine tools, cataloging systems, Optical Character Recognition and display technologies used largely "off-the-shelf" available open source technologies, with minimal additional configuration required.

Incorporation of languages such as Syriac, Arabic, and Aramaic posed significant new hurdles. For instance, works with titles, abstracts and descriptions in Syriac must have this information entered into cataloging records in the native language, a transliterated form, and an

English-translated form. Although care had been made early on to support UTF-8 encodings in RUCore, this was no guarantee that the software would indeed correctly handle the full character set. The only way to be certain would be to test the software with real data.

In practice, additional code was required to permit the printing of Syriac script on RUCore search interfaces and record displays. Some additional coding was required to permit users fluent in Syriac and Middle Eastern languages to enter search terms in the native scripts, and have the search software parse our records and return accurate results. Data entry of these records was greatly enhanced by the availability of spreadsheets and basic electronic catalog records of Beth Mardutho’s collection items, permitting metadata entry to commence by direct cutting and pasting into the OpenWMS cataloging interface all of the titles and abstracts written in Syriac and other languages.

Modification of the search interface and indexing engine also became necessary. Previously, RUCore had used an open source search engine software known as Amberfish to index resources and provide search results to end users. Our tests with indexing characters which contained diacritics, double-byte characters and non-Latin character sets proved discouraging, and it became evident that multi-language support with Amberfish was less than desirable. As a result, the development team searched for alternatives, preferably of a similar open source nature, and we found that a great deal of enthusiasm existed for SOLR/Lucene, which provided not only multilingual support, but greater efficiency and processing speed, while using fewer computing resources (Arnold, 2012). As a result, testing began with an upgraded SOLR-equipped version for the software. After some configuration changes to our MySQL database server to ensure that Unicode character sets were supported, the software quite easily supported the entry of Arabic characters, and could effectively index and search these entries.

A greater challenge involved how to handle the varied degrees of support for Syriac on end-user computing systems. Official support for Syriac is available on Microsoft Windows-based computing systems, beginning with Windows XP. Microsoft provides a Syriac phonetic keyboard layout, but the necessary software extensions may not be installed by default, and require additional setup (“Accents, Symbols and Foreign Scripts,” n.d.). By contrast, as yet the Mac OS X operating system on Apple-based hardware does not provide a built-in Syriac keyboard layout, however third-party freeware add-ons exist. On both platforms, display of Syriac script is possible through the operating systems’ native use of UTF-8 encoding, but in some cases the necessary font required to translate the encodings into actual script may be missing.

In order to assist users, links were provided directly on the search interface to enable the installation of fonts for correct Syriac display. Users are advised that if they experience trouble seeing certain text in the search results and record displays, they may download and install the needed fonts to remedy the problem.

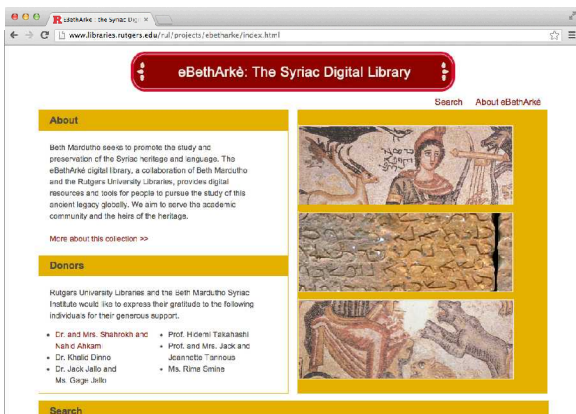


Figure 2: Primary Landing page

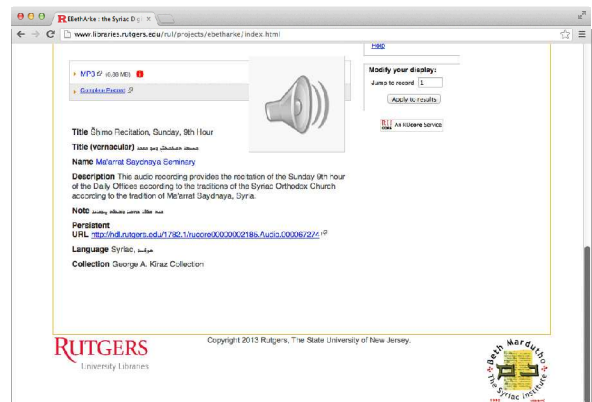


Figure 3: Display of a digital library object, with Syriac script.

## Current Status

The website and associated electronic library interface can be publicly accessed at <http://www.libraries.rutgers.edu/rul/projects/ebetharke>. The initial 110 objects include two photograph insets from a book on Syriac texts and literature; a scanned Arabic periodical titled *Tārīkh Sūriyyah* (native Arabic title *تاريخ سورية، المجلد السادس*); a 60-page description of the Syriac language in Latin from CE 1573; a letter from Elias II (d.1932), the last Syriac patriarch known to reside in Turkey; and an 1898 Syriac periodical from Beirut. Additionally, two multimedia objects are demonstrated: one sound recording of a Sunday 9<sup>th</sup> Hour *Šhimo Recitation* (native title “ܫܝܡܘ ܩܘܨܐܬܐ ܨܝܢܝܘܬܐ ܨܝܢܝܘܬܐ ܨܝܢܝܘܬܐ”), and video of an interview with Fr. Shimon Can of St. Mark’s Monastery, Jerusalem, one of the few remaining scribes of the Syriac language.

## The Road Ahead

The technical solutions discussed earlier solve the issue of native Syriac support on desktop and laptop computing systems. Mobile devices, which are becoming ever more prevalent, pose a greater challenge due to the limited nature of their operating systems, and at times limited ability to modify and extend the operating systems. As of this writing, no native display or data entry method exists for Syriac on

iOS devices (such as iPhones and iPads), nor can a third party option be used due to the closed nature of the system. Similarly, the Android mobile operating system lacks support for Syriac, and while theoretically feasible, no third-party extensions exist for this purpose.

Notably missing as well is an available Optical Character Recognition engine for the Syriac language and character sets. While the repository platform can correctly search for terms cataloged in Syriac using native characters, these must all be entered manually as part of the catalog process. At present, there is no known method for digitally recognizing written Syriac text, and making such text automatically searchable. Overcoming this challenge will require substantial development resources, in terms of staffing and monetary commitments. Other minor improvements to consider in a future development roadmap include shoring up the controlled vocabulary used for cataloging and metadata entry. Feedback, including from evaluators of this article has suggested that more authoritative databases should be integrated into the platform.

Funding for Beth Mardutho will also continue to be an ongoing challenge. Currently the institute exists primarily through the financial support of charitable donations. As such, staffing and resources at the institute are limited. This will in turn affect the speed and volume with which collection items are digitized and added to the digital library collection. It is hoped, however, that the inclusion of Beth Mardutho into RUcore, and the ongoing presence of an established digital library architecture in a high-profile research institution, will raise the stature and awareness of the institute and its efforts among potential donors who would be willing to provide ongoing financial support to further Syriac teachings and heritage. Such possibilities bring hope and optimism that the institute will thrive going forward, and that Syriac literature, language and artforms will continue to live in the online space for years to come.

## ACKNOWLEDGEMENTS

The execution of this project would not have been possible, of course, without the primary partners in this endeavor. George Kiraz, founder of Gorgias Press and the Beth Mardutho Syriac Research institute, has led a long and tireless effort in ensuring that the language lives on in the academic, cultural and digital realms. Grace Agnew, Associate University Librarian for Integrated Information Systems at Rutgers University Libraries, is the driving force behind RUcore, and its master architect. There are also a stable of developers, policy makers and structural decision makers. A non-inclusive list of these talented individuals, with apologies for anyone I may have left off: Kalaivani Anathan, Jie Geng, Dave Hoover, Dr Ronald Jantz, Linda Langschieid, Rhonda Marker, Chad Mills, Dr. Jeffery Triggs and Yang Yu.

## REFERENCES

- [1] "Accents, Symbols and Foreign Scripts" (n.d.), Syriac/Aramaic/Modern Assyrian. From Penn State *Teaching and Learning with Technology*. Last accessed July 19, 2016 from <http://symbolcodes.tlt.psu.edu/bylanguage/syriac.html>.
- [2] Agnew, G; Maher, CA, and Mills, C (2010). "VMCAlytic: Developing a Collaborative Video Analysis Tool for Education." *Proceedings of the 43rd Hawaii International Conference on System Sciences – 2010*. doi: 10.1109/HICSS.2010.438
- [3] Agnew, G. and Yu, Y (2007, December 17). "The Rutgers Workflow Management System: Migrating a Digital Object Management Utility to Open Source." *The Code4Lib Journal*, Iss 1. 2007-12-17. Last accessed July 19, 2016 from <http://dx.doi.org/doi:10.7282/T3JM280B>
- [4] Arnold, S. (n.d). LUCENE/SOLR: Now Ready for the Big Leagues. *Online*, 36(6), 40-43.
- [5] Beard, I; et. al. (n.d.). "Digital Curation: Current Standards and Documentation." From Page2Pixel: The Digital Curation Blog at Rutgers University Libraries. Last accessed July 19, 2016 from <http://page2pixel.rutgers.edu/standards>.
- [6] Gibson, N. P., Michelson, D. A., & Schwartz, D. L. (2016). From manuscript catalogues to a handbook of Syriac literature: Modeling an infrastructure for Syriaca.org. March 3, 2016; Available from: arXiv, Ipswich, MA. Last accessed July 19, 2016 from <http://arxiv.org/pdf/1603.01207v1.pdf>
- [7] HEAL, K. S. (2012). CORPORA, ELIBRARIES AND DATABASES: LOCATING SYRIAC STUDIES IN THE 21ST CENTURY. *Hugoye: Journal Of Syriac Studies*, 1565.
- [8] Jantz, R and Giarlo, M (2005). "Digital Preservation: Architecture and Technology for Trusted Digital Repositories." *D-Lib Magazine*, June 2005. Last accessed July 19, 2016 from <http://dx.doi.org/doi:10.7282/T3DV1H8M>
- [9] Jeng, J (2008, November 4). "Evaluation of the New Jersey Digital Highway." *Information Technology and Libraries*, Vol 27, No. 4 (2008). doi: 10.6017/ital.v27i4.3238
- [10] Kahn, R. and Willensky, R. (13 May 1995). *A Framework for Distributed Digital Object Services*. cnri.dlib/tn95-01. Last accessed July 19, 2016 from <http://www.cnri.reston.va.us/k-w.html>
- [11] Kiraz, George (2001). "EbethArké: The Syriac Digital Library, First Report." *Hugoye: Journal of Syriac Studies*, Vol. 4.2, 269.–271
- [12] Schild, T. (n.d.) "Syriac UTF-8 Character Set: page with code points U+0000 to U+00FF" UTF-8 encoding table and Unicode characters. Last accessed July 19, 2016 from <http://www.utf8-chartable.de>