Digital Library Solutions

Iqura Technologies

INTRODUCTION

Historically, a library has served as the cornerstone of mankind's endeavor to learn and disseminate knowledge. The library acts as a central repository of our combined learning with an aim of making it freely accessible to society at large.

Libraries today are reinventing themselves to meet the evolving demands of our increasingly networked world and to reach out to a larger audience. From being a building that houses paper books, the library is metamorphosing into an online storehouse of a vast range of digital content. This library is far more accessible to audiences, highly available and provides a whole new range of tools for the seeker of knowledge. Moreover, this virtual library is now available at your fingertips.

Digital library technologies and practices have evolved and developed greatly in the recent past, to a point where they are now within easy reach for most institutions that manage content such as libraries, museums and educational institutions.

Iqura has been making definitive contributions to the evolution of digital libraries and development of best practices. This document presents our approach to the development of a best-of-breed digital library.

Iqura is a leading IT services provider, with an excellent track record of over six years in delivering quality solutions to global customers. We provide technology services across major practices with extensive competency in media & publishing, financial services, energy & utilities and real estate domains.

Our aspiration is to go beyond the traditional definitions of consulting, by innovating to bring you solutions that ***dramatically improve*** the efficiencies in your business. We understand the new economy, the advantages and the pitfalls, and are ideally positioned to architect your company's business strategy in this changing business environment.

At the core of our activities is a management team with the highest levels of commitment to the customer, industry knowledge, technical expertise and strategic insight of technology trends. We have put together a team that can speak to you on your terms and in language you understand; without the hype and the jargon; a team that will ***define a solution that works and works for you***.

To learn more about how we can help your digital library initiatives, mail us at info@Iqura.com.

WHAT IS A DIGITAL LIBRARY?

In a real-world library, a patron normally travels to a physical location during certain fixed hours to locate needed information. Individuals' expectations for access to information from a library have increased dramatically with the emergence of the Internet. Patrons are no longer satisfied with the speed of access wherein they have to visit the library, physically search for an item, check for availability and/or wait to receive it. Increasingly, they expect instant access to the contents of the library and an ability to search for the information they need at any time, from any location and through different devices.

This is the need that digital libraries aim to fulfill.

In essence, a digital library presents a very large collection of information in an easy-to-use coherent way. However, an emphasis on content simply being in digital format is too limiting, the objective has to be develop the information systems necessary to provide structures which enable access to a coherent collection of material, in whichever format they might be, including in digital formats.

As digital libraries evolve, they begin to realize the vast opportunities available by materials that are in digital format and ways to exploit these opportunities. Thus the digital library gets strengthened and adds value by its ability to integrate materials in digital formats that may not be as well-represented, easy-to-access or effectively usable in traditional library environments. For example, the ability to intersperse different formats such as multimedia, statistical data, maps and geospatial data, from different sources within a single collection and their presentation as a coherent whole.

The benefits that digital libraries bring to a library patron include:

- **Access** to the varied collections of libraries across geographies through automated catalogs.
- Ability to **locate** both digitized as well as physical versions of scholarly text and books.
- Ability to **search the content** in an efficient manner whether within the library, through the Internet or through other commercial databases.
- Ability to **instantly access** the relevant content from the search results by clicking through to the digitized content or find additional items of relevant interest.
- Ability to **conduct more detailed research** by being able to save search results, as well as use different criteria and parameters to narrow down their search results.

The major advantage is that all the above benefits are available to a user from her desktop directly or increasingly from other web-enabled devices such as personal digital assistants or cellular phones.

Very often, the term “digital library” is used to describe any multi-media management system that contains digitized information. However, these systems may not necessarily deliver true library application functionality. It is also necessary that digital libraries co-exist and interoperate with existing library infrastructure. Therefore, the development of a digital library system must be implemented such that it works with existing library catalogs and incorporates industry standards, protocols and formats. As digital libraries have evolved, there are now well-defined as well as emerging standards on how the different components of a digital library should fit together and the mechanisms for information capture, encoding, storage, delivery and archival.

Libraries digitization and automation efforts must also take into account the issue of content ownership and their intellectual property rights management. The Internet is increasingly used to buy, sell, or license the use of content, be it documents, images, video or other forms of copyrighted content. This then imposes the responsibility on the library to protect that content from unauthorized use and or define the access to that content for its patrons.

IQURA’S DIGITAL LIBRARY SOLUTION

Iqura’s digital library solution is an initiative to provide **a standards-based, open-source, cross-platform system** for the implementation of a digital library **that can support multiple languages, has extensive search capabilities, and provides administrative mechanisms for management of the library**. The solution provides for an end-to-end development from digitization to deployment within a time-bound manner; and one that allows for emerging standards and new features to be incorporated as the library grows.

DIGITIZATION

The first step in any digital library solution requires the digitization of content. Iqura provides digitization facilities with high levels of accuracy with multiple levels of error-checking.

CONTENT AND METADATA ENCODING

Iqura’s solution encodes the digitized content using an XML standard called “Text Encoding Initiative” or TEI. The source content can be languages other than English as well, since it is encoded using the UNICODE standard. Thus, the solution is equally suitable for libraries with content in different languages.

The files comprising the digital version of a manuscript or other scholarly text would likely include images of varying formats and resolutions (tiff master, medium resolution jpeg, thumbnail gif) and/or a structured text transcription (TEI/XML) of the manuscript.

The key to locating, using, and preserving digital content is metadata, or structured contextual information about digital objects and collections. Many digitization efforts have been unsuccessful due to inadequate metadata. There are three different types of metadata, all essential to ensure the usability and preservation of the collection over time. These are descriptive, structural and administrative metadata.

A good encoding of metadata will ensure:

- **Full inventorying** of the resources or files that make up the content of a digital object.
- Specification of **how the content files all fit together** into a coherent whole expressing descriptive and administrative metadata pertaining to the content.
- **Encoding of information** on how the individual components that make up the object relate to each other, including the order in which they should be presented to the user. For example, how the still image files that comprise a digitized version of a print volume should be ordered.

The metadata necessary for a complete description of the books is encoded in another XML standard namely, “Metadata Encoding and Transmission Standard”. METS is a standard for packaging descriptive, administrative and structural metadata. Leading academic and research libraries are citing METS as an important standard for digital library interoperability, and it is emerging towards becoming the de-facto standard.

METS is a newly emergent standard designed to encode all varieties of metadata necessary for a complete description of digital objects within an electronic library. Such objects may take the form of electronic texts, still images, digitized video, sound files or more interactive material such as VRML virtual environments.

It allows for metadata that adheres to existing standards (such as Dublin Core and MARC) to be embedded in a METS record, or stored outside the METS record and referenced. In some ways, METS is therefore not a metadata standard but rather a wrapper for associating existing metadata of various types within a single object, document, or collection structure.

Both the TEI and METS standards used in Iqura’s solution are expressed in XML, which is an open, extensible and interoperable (platform independent) way for data interchange.

BUILDING A DIGITAL REPOSITORY

Our digital library solution is built upon the Greenstone platform. Greenstone is a suite of software for building and distributing digital library collections. Greenstone is produced by the New Zealand Digital Library Project at the University of Waikato, and developed and distributed in cooperation with UNESCO and the Human Info NGO. It is open-source, multilingual software, issued under the terms of the GNU General Public License.

Iqura has built extensively upon the Greenstone platform to offer the flexibility and feature-set expected by demanding users of today. We have extensively modified and enhanced the base platform to develop advanced features not supported by Greenstone. In addition, we have built the interfaces necessary for Greenstone to understand the TEI/METS encoding which are imperative for a standards-based library.

LIBRARY PRESENTATION

The contents of the digital collection of the library may vary from text such as books, manuscripts, and plays to other forms such as music, images, etc. The solution offers flexibility to handle the different types of content and their presentation. For example, books can be presented as formatted, searchable electronic text along with scanned images of each page.

For maximum accessibility the texts are presented as HTML, and adhere to best practices and appropriate standards for text conversion, content management, data archiving and web delivery.

Adherence to "Web Accessibility Initiative (WAI)" guidelines

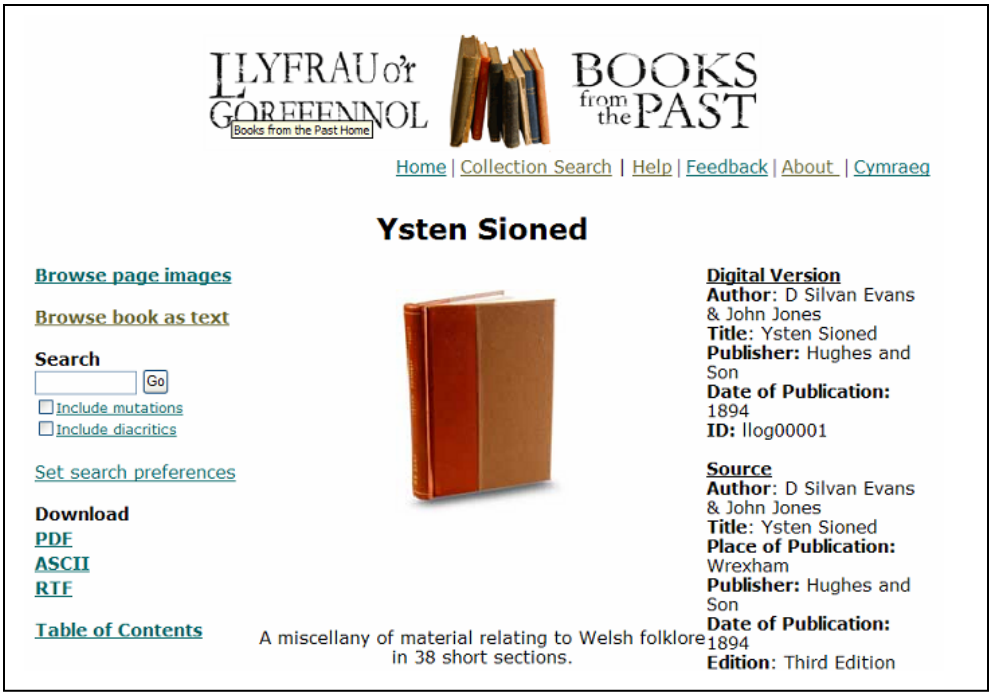
The library website conforms to appropriate Web Accessibility Initiative (WAI) guidelines. In particular the site is accessible to visually impaired users, to those with low speed (36kbit/s) connections and to those who use screen and/or text readers.

Conformance to HTML standards

The library website's content is presented in HTML, which in turn is validated against 'HTML 4.01 transitional, W3C WAI level A' specification that conforms to W3C Recommendations.

Quality standards

The books are digitized, encoded and reproduced on the web with 99.99% accuracy of electronic text on a character basis. The books are faithfully reproduced in terms of content, format and annotations



LIBRARY NAVIGATION

The library supports multiple languages for site navigation so as to address a larger number of potential users. The following section describes the various navigation aids provided for a bilingual digital library solution in Welsh and English.

Bilingual Navigation

The site can be browsed in either Welsh or English. The content remains in the source language (Welsh or English).

Character Encoding

Special attention was given to Welsh diacritics and original orthography.

Site Navigation

Books can be browsed page by page (exactly matching the source books) either as text or as an image. A provision to switch to either text or image mode of browsing while keeping the browse position unchanged was provided.

Table of contents

A short Table of Contents has been provided on the left navigation bar, with only main sections/chapters listed.

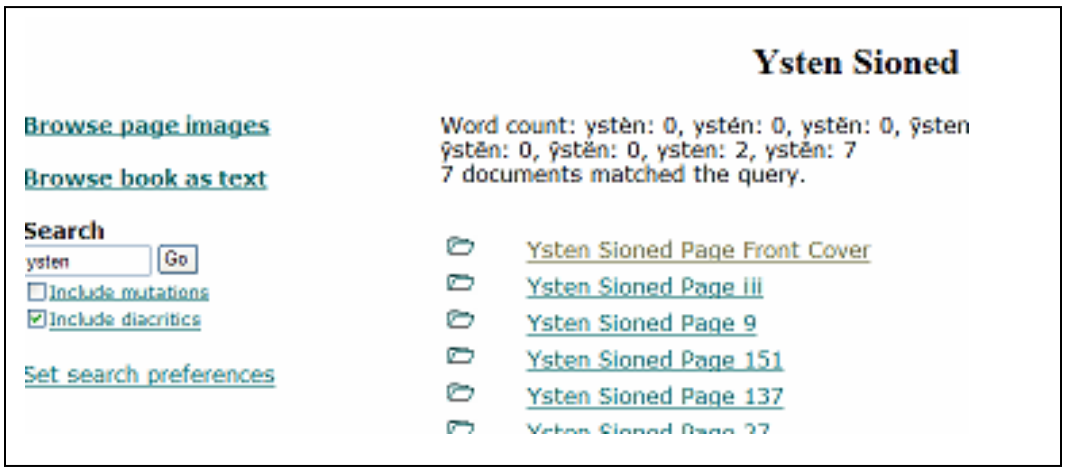
A Full TOC appears on a separate page, listing main sections and sub-sections. It is possible to jump directly to a specific page from the TOC.

Navigation aids

Navigations aids were provided to navigate the book, page-by-page using Next, Previous, First and Last options. The reader can also directly go to any page by keying in the page number.

LIBRARY SEARCH CAPABILITIES

One of the key advantages offered by a digital library is the ability to search a digital collection for specific information resources. Therefore, the quality of the search engine and the flexibility provided in searching become very important. Iqura has enhanced the search capabilities extensively, over and above the capabilities provided by the Greenstone platform. Iqura's digital library solution has extensive search facilities with specific enhancements for searching non-English text.



The search facilities include:

- **Searching entire books or collections** – it is possible to search for a specific phrase, keyword, etc throughout the collection or within a specific book.
- **Searching with exact diacritics** – A diacritic is a mark added to a letter or symbol indicating a change in its usual pronunciation, e.g. à, ê, ô. It is possible to search for a specific keyword with the diacritic marks included.
- **Searching with mutations** - Mutations are instances where the initial letter of a word changes depending on its grammatical context. For example, in Welsh, the word for "stone" is "carreg", but "the stone" is "y garreg" (soft mutation), "my stone" is "fy ngharreg" (nasal mutation) and "her stone" is "ei charreg" (aspirate mutation). It is possible to search for a specific phrase or keyword across it's various mutations.
- **Advanced filter options** were provided to narrow down search results. These include search by author, title, language, publisher, period of publication, subject and/or just plain text search.
- **Boolean search** using operators like AND, OR and NOT is also possible.
- **Saving searches** throughout your session on the website is possible through the search preferences settings which can be customized to individual requirements.

LIBRARY ADMINISTRATION

Any library needs an Administrator to maintain it and so it is with digital libraries.

This administrative interface permits easy review and editing of METS and TIE XML source files. This interface allows a side-by-side viewing of XML and the corresponding scanned page images. Other important and pertinent features of the administrative interface are:

- **Secured access** to the Administrative interface with username and password
- Supports **multiple users having different editorial roles** and hence different access rights
- One-step **easy addition, modification and deletion** of books from the collection.
- Ease of modification of source XML files, with **in-built simple editor**
- **Built-in validation tool** for TEI DTD and METS schema.

SUMMARY

In this introduction, we have outlined our digital library solution to you within the overall context of our media and publishing practice area. It is our belief that these services would provide significant value towards meeting your requirements.

Iqura has deep document management and publishing skills built over a period of six years through multiple large turnkey solutions developed for enterprise clients. Our expertise ranges from custom-built publishing solutions for companies such as Cyber India Online (CIOL) to platform driven solutions based on Greenstone, SharePoint Portal Server, Site Server, etc. We also have extensive application development experience for document management and publishing such as data entry systems, large automated doc-to-PDF solutions and standards-driven (such as XML, SGML and TEI) systems.

Iqura will undertake to assess your requirements, evaluate possible options, and recommend potential solutions and implement the entire development process with you, our credentials for this assignment have been outlined in the document.

Please do feel free to contact us for further information on these and/or any of our other services.

We look forward to the opportunity of working with you for your development and consulting requirements.

Iqura Technologies Pvt. Ltd.
5/1-87, Hosur Road, Madiwala, Bangalore 560068, INDIA
Telephone: +91-80-25529209/210. Fax: +91-80-25529597
www.iqura.com, info@iqura.com

NOTICES
This document contains confidential information about Iqura's services, methodologies, and pricing and internal processes. This document should be circulated only to your employees or customers on a need-to-know basis.

© Iqura™ Technologies Pvt. Ltd., 2006. All rights reserved.