BIBLIOTHECA ALEXANDRINA
مكتبة الإسكندرية

International School of Information Science

International School of Information Science

BIBLIOTHECA ALEXANDRINA
مكتبة الإسكندرية

International School of Information Science

Alexandria
January 2005

© Bibliotheca Alexandrina, 2005

All rights reserved by the Bibliotheca Alexandrina, International School of Information Science (ISIS). However, for non-commercial purposes, this document may be freely reviewed, abstracted, reproduced, translated in part or in whole, stored or introduced in a retrieval system or transmitted in any form or by any means, electronic, mechanic, photocopying, recording or otherwise without the prior permission of the Bibliotheca Alexandrina International School of Information Science (ISIS) provided that the source is acknowledged/credited.

Contents

Introduction 5

Mission 5

Aims and Objectives 6

Projects 6

The Internet Archive 6

Million Book 8

The Digital Library of the Modern History of Egypt 9

Gamal Abdel Nasser Collection 9

Description de l'Egypte 11

Pioneering a new business model for Arab electronic publishing 12

Online Access to Consolidated Information on Serials (OACIS) for the Middle East 13

Revival of L'Institut d'Egypte 13

My Book: Digital and Printed 14

Hole in the Wall 15

The Development Gateway of the World Bank 15

Virtual Visits 16

The Supercourse 17

Universal Networking Language (UNL) 17

International School of Information Science

Introduction

In order to maximize creativity and reduce the delay between an innovation and its full implementation within the Bibliotheca Alexandrina (BA), BA founded a research institute named the International School of Information Science (ISIS).

Guided by BA goals to:

- preserve the heritage for future generations in digital form; and
- provide universal access to human knowledge

ISIS acts as an incubator for digital and technological projects, promoting and nurturing innovations which view the encompassing aims of BA as an end goal.

Mission

ISIS was founded to initiate, carry out and promote research and development of activities and projects related to building a universal knowledge center. The Institute is viewed as an incubator of IT projects that will ultimately contribute to the knowledge capacity of Egypt and the world.

Aims and Objectives

ISIS aims to be the focal point where scholars from around the globe will group to study and develop new technologies serving BA's mission of becoming a true library for the digital age. Using state-of-the-art technology, ISIS aims to join hands with other organizations, institutions and IT centers around the world to research and implement pioneering digital ideas for the benefit of the international knowledge community. In fact, the Institute has already created partnerships and adopted a number of major projects in accordance with BA mission.

Projects

The structure of ISIS is based on projects where staff are either recruited from the Information and Communication Technology (ICT) Department or hired specifically for a particular project. Once the project is in the production phase, it joins the mainstream within the concerned BA department with the technical support provided by the ICT Department. Some of the past and ongoing projects are outlined below.

The Internet Archive

http://archive.bibalex.org

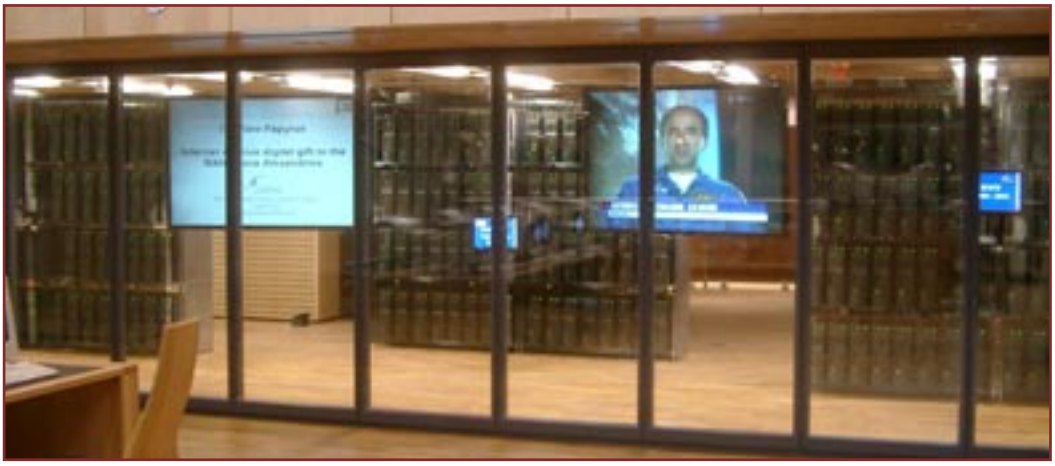
In January 2002, the Internet Archive in San Francisco donated a copy of the Internet Archive to BA. The objective was to establish the second center worldwide holding a full copy of the Internet Archive. The Internet Archive is a complete snapshot of all web pages on every website since 1996 till today. Since the average lifetime of a page on the Internet is 100 days, this snapshot is retaken every two months. The archive at BA (http://archive.bibalex.org) includes 10 billion web pages from 1996–2001; 2000 hours of Egyptian and US television broadcast and 1000 archival films. It represents 100 terabytes of data stored on 200 computers. The archive has been fully operational and the collection is accessible through the BA website and synchronization with the Internet Archive in San Francisco is carried out regularly.

A special design for the machines used for storing the archives was applied at BA, and a prototype system,

which has been successfully used in other projects, was installed. An agreement for building the second generation machines for web archiving, the Petabox, is currently underway. The Petabox is a machine designed to safely store and process one petabyte (a million gigabytes) of data. The machine features low power consumption, multi-operating systems, easy maintenance and software to automate mirroring. The new machines will be installed holding the web collections of 2002, 2003 and 2004 as well as collections of video, audio and text. New machines for the 2005 collection will be designed and manufactured locally.

The WayBack Machine allowing people to surf the web as it was

The Internet Archive displaying an archived TV broadcast



The Digital Laboratory

In line with the BA overall mission of building a universal digital library, and since many of the projects hosted by ISIS require specialized digital services, ISIS has built its own digital laboratory equipped with state-of-the-art technologies. The laboratory is used for digitizing various media including slides in multi-formats, negatives, books, manuscripts, pictures, maps, audio and video, and is equipped with the necessary tools used for indexing, archiving and management, thus automating the entire workflow. The building of the laboratory was an essential starting point in digitizing BA collections and collections of other international libraries that are interested in pursuing the goal of "Universal Access to Human Knowledge". Many of the projects outlined below are highly dependent on the Digital Laboratory to achieve their goals.

Cruse Scanner

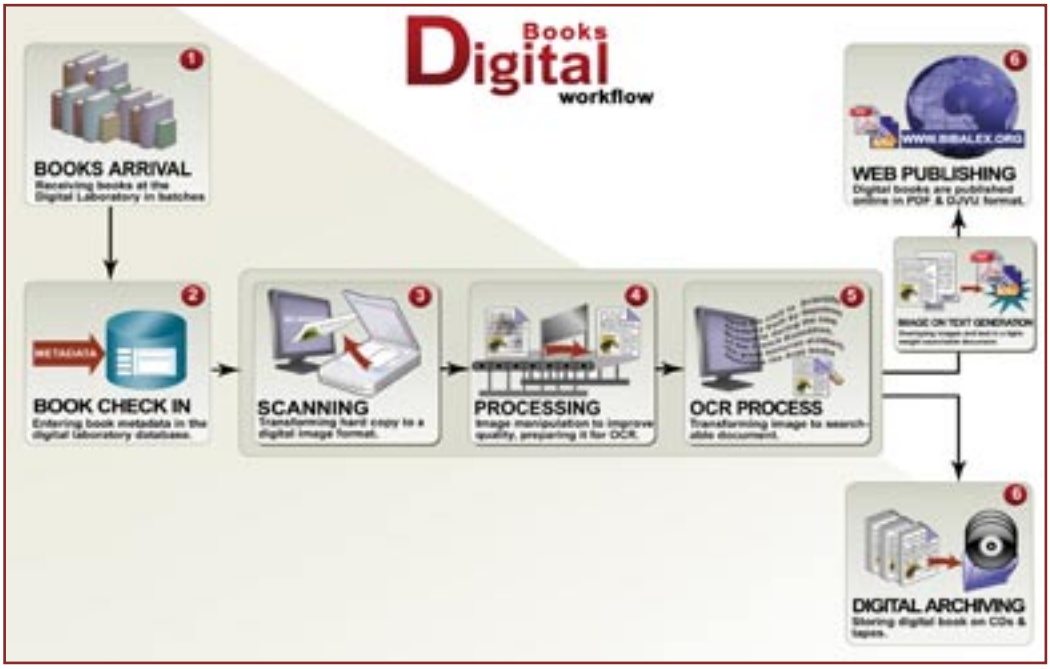
Million Book

Minolta PS7000 scanner for scanning face-up originals

Initiated by Carnegie Mellon University, the primary long-term objective is to transfer all books into digital format, in partnerships with other scanning centers internationally, in order to create a Universal Digital Library, which will foster creativity and free access to all knowledge. BA and its partners (including China, India and USA) have been working to demonstrate the project's feasibility by digitizing one million books within three years and publishing them as a searchable collection on the Internet. All project partners will contribute content to ensure that the collection is extensive, diverse and multilingual. BA is expected to take the lead in scanning and digitizing 75,000 Arabic books in a three-year time frame. The target is to scan and process 5000 pages/day/scanner (2 shifts), which will lead to approximately 25,000 books/year. Moreover, the project will provide a test-bed that will support research on improved scanning techniques, optical character recognition (OCR), and indexing.

A database for the books, metadata and status has been designed and implemented and standards have been set for the process of digitization in order to improve quality in the scanning, processing and OCR phases. The complete

Digital Lab workflow



cycle of the workflow for producing digital books has been automated and integrated with the Library Information System. The database was further expanded to accomodate scanned images and slides. In addition, research was carried out in co-operation with Arabic OCR producers in order to achieve efficient, high quality recognition for mass OCR production. Preliminary experiments for supporting the Arabic language with Image-on-Text on DjVu format have also been conducted and produced satisfactory results. Work is underway to support the Arabic language with Image-on-Text on PDF format.

Inside the Digital Lab

The Digital Library of the Modern History of Egypt

This integrated digital library will include the collection of specialized libraries belonging to eminent Egyptian politicians, authors and historians as well as content from all over the world related to the modern history of Egypt encompassing the past 200 years. The project aims to scan, catalog, index and OCR the collection and present it in a searchable form to its users, and thus benefit from the efforts exerted in the Million Book project. Agreements are being negotiated with several individuals, foundations and libraries to provide the content of specialized collections for scanning. The Nasser collection and *Description de l'Egypte* projects are outcomes of such agreements.

Gamal Abdel Nasser Collection

www.nasser.org

In cooperation with the Nasser Foundation, BA has digitized the collection of Egyptian President Gamal Abdel Nasser and published it through an integrated searchable web-based system, which is mainly intended for research and documentation purposes. The collection is continuously expanding with new entries and includes the following items:

- Documents published by the Public Records Office, London, United Kingdom, between the periods 1919 to 1995 (around 1540 topics in more than 52,000 documents);
- Documents published by the United States Department of State (16 volumes of 7,965 documents in 13,939 pages) during the period 1951–1978;

Home page of Gamal Abdel Nasser website

- Over 1300 speeches given by the President during the period 1952–1970, in both audio and printed format;
- Over 51,000 photos and 1000 portraits of the President covering more that 6000 events during the period 1930–1985;
- More than 1,230 videos (over 50 hours) taken during the period 1948–1970;
- A complete archive of the articles published in the newspapers about the President and the 1952 revolution;
- The decrees issued by the Revolutionary Command Council (RCC);
- The daily news of the President;
- Minutes of the Central Committee for Arab Socialist Union (ASU);
- The President's personal correspondence and other relevant documents in his own handwriting (596 documents under 145 topics);
- A complete archive of the special weekly column "Bisaraha" by the Egyptian writer and journalist Mr. Mohammed Hassanein Haikal documenting all major events during the period 1957–1994 in simple writing for the general public;
- Books written by and about Abdel Nasser;
- A number of items representing the effect of Nasser on the culture of the region. The collection includes 1431 national songs, 138 poems, 50 rare stamps, 125 caricatures, 34 plastic arts illustrations, and 38 commemorative and circulating coins.

The collection was scanned, cataloged, indexed and subjected to OCR. A database has been constructed for each topic storing each document's text along with its meta data. The entire collection was presented in a web based interface with full-text Arabic and English search in both meta data and content. The website (www.nasser.org) facilitates browsing of the collection through displaying the various items as well as providing full text and morphological searching, presenting rich information for both researchers and the public. In fact, the major strength of the Nasser project lies in the integration of all the different information sources and media under a single interface and in a fully searchable form. Hence, a researcher may find all possible documents, pictures, videos, stamps, articles, etc., pertaining to his/her field of interest by performing a single search.

DESCRIPTION DE L'EGYPTE

*Description de l'Egy*pte was the outcome of the collaboration of more than 150 prominent scholars and scientists who accompanied Napoleon in 1798, and some 2000 talented artists and technicians. For over 20 years, they systematically examined almost every aspect of contemporary and ancient Egyptian civilization, producing 20 volumes of text and plates of unmatched accuracy and detail. Historically these engravings became the most comprehensive record and inventory of Egypt's land and monuments.

Description de l'Egypte, virtual browser main interface

Drawing of Muhammad Ali lounging at his royal abode, "Description de l'Egypte"

This valuable collection containing images related to antiquities, natural history, and the modern states of Egypt has been fully digitized and integrated on a virtual browser with the objective of preserving it and making it publicly accessible. The collection includes eleven volumes of plates owned by BA, as well as ten volumes of text, a contribution from *Institut d'Egypte*. A tool was developed to publish books in the standard extended markup language (XML) format where books may be browsed by a virtual browser or touch screen. The project is composed of three stages. In the first stage, the application is provided on DVDs in two versions for the public and for researchers. During the second stage, the relation between the text and images were established and rendered in a searchable form. In the final stage, the whole collection is to be available in an integrated searchable form on the web.

Pioneering a New Business Model for Arab Electronic Publishing

The main objective of this project is to design and implement an innovative business model for the e-book supply chain for Arabic books and to test the standards and protocols through building a prototype of Print-On-Demand (POD). Electronic publishing and POD have created fundamental changes in the production, management, dissemination, and use of all types of information. Authors are able to have their works online where the information may be easily accessed, browsed, annotated, critiqued, downloaded, and shared. The result is significant changes to the linear path of writing, refereeing, and reviewing of publications as all these functions can be performed concurrently.

BA is partner with one of the Egyptian publishers (Dar el Sherouk) to implement a new model using a POD system within BA, where BA will act as a retailer (POD outlet). This will enable testing all the necessary requirements for generalization of the concept in a later phase. Upon the accomplishment and operation of this prototype, similar units will be installed in other Arab countries.

Specialized font libraries for OCRing of the books from Dar el Sherouk have been developed to OCR the entire

collection. Over 1000 of their published books have already been scanned and OCRed. Moreover, an analysis study of the current electronic format applied by the publisher is underway to determine its applicability in electronic publishing.

Online Access to Consolidated Information on Serials (OACIS) for the Middle East

Initiated by Yale University Library, OACIS proposes the creation of a publicly and freely available electronic union list of serials and journals from or about the Middle East. The mission of OACIS is to improve access to Middle Eastern serials in libraries in the United States, Europe, and the Middle East and to make scholarly literature from, and about, the Middle East widely and easily available to scholars around the world.

The OACIS project official website

The OACIS system was launched in November 2003 and currently comprises 12 partners, 42 languages (with the top collections in Arabic, Persian and Turkish) and 25,000 title records. A partner since August 2004, BA is contributing to the project through:

- Enriching the OACIS catalog with the records available in the BA catalog with quarterly updates;
- Digitizing serials and journals and implementing a content - retrieval web interface for the digitized material;
- Designing and implementing an Arabic web interface for the collection; and
- Setting up a mirror site of the system at BA.

www.library.yale.edu/oacis/

Revival of L'Institut d'Egypte

BA is taking the initiative in reviving the organization of *Institut d'Egypte* built in Cairo by Napoleon Bonaparte over 200 years ago. The *Institut* first scholars were in charge of the research, study and publication of physical, industrial and historical facts about Egypt, publishing findings that stemmed from their activities as members of that body. They were later to produce the *Description de l'Egypte*. Eventually, the *Institut d'Egypte* became the focal point for scholarly

Institut d'Egypte built by Napoleon over 200 years ago

work and intellectual pursuit in Egypt, and provided both actual space and structure for scholarly discourse. It is also recognized as the oldest functioning academy of sciences and arts outside Europe.

In order to preserve the *Institut d'Egypte* collection and make it accessible to the public, the Bibliotheca Alexandrina has suggested nine projects for its revival, among which is a project of digitizing its entire collection. The effort has started with the digitization of the 10 volumes of *Description de l'Egypte.* Other special collections were digitized such as the complete works of Voltaire (69 volumes), *Des Mille Nuits et Une Nuit* (16 volumes), and *Geographie Universelle* (15 volumes). Eventually, the whole library of the *Institut* will be digitized and made available to the public.

My Book: Digital and Printed

http://archive.bibalex.org/mybook

This project is designed to enable children to relate to both printed and digital information in a seamless fashion, and to bring the marvels of the digital age to the poorest parts of the community. The project teaches children that written words can be transformed from digital format to printed format and then to a bound book and vice-versa.

My Book: Digital & Printed process cycle

They are attracted to the project booths where they are allowed to print and bind books, and retain them too. A portal has been implemented with over 250 digitized Arabic children books and a database has been designed and implemented to hold their meta data. (visit http://archive.bibalex.org/mybook)

Bookmobile unit at one of the schools

A stationary unit has been installed near the entrance of the Children's Library and is operational since October 2003. Additionally, a Bookmobile unit has been touring more than 80 schools since November 2003.

Hole in the Wall

Based on the highly innovative and successful experiments conducted in India, BA intends to bring children of the community of Alexandria in direct contact with the cyber world and its infinite possibilities for education and learning. The "Hole in the Wall" project is based on the hypothesis that children learn to use the computer and access the Internet independently through self-learning.

The idea was initiated in India by Hole In the Wall Education Limited (HIWEL), where free Internet street kiosks were provided to children in rural areas who were unfamiliar with computers and the Internet. It was observed that, even in the absence of any direct input, mere curiosity led groups of children to explore, which resulted in learning. This led to the belief that any learning environment that provides an adequate level of curiosity and motivation can promote learning in groups of children, with no intervention from adults. It was further proven that it is possible to install a computer, connect it to the Internet, and keep it in working condition in any outdoor environment despite heat or dust, etc.

HIWEL free Internet street kiosks

A portal holding some interesting sites for children, with games, stories and educational tools has been built. A prototype of one kiosk has already been set inside BA and another in the BA plaza. In cooperation with the Governorate of Alexandria, 30 other locations are being selected to establish the kiosks. A committee comprising the best social researchers in Egypt are designing the experiments to be conducted and analyzed.

A prototype of one kiosk has already been set inside BA

The Development Gateway of the World Bank

The Development Gateway is an interactive portal for information and knowledge sharing in sustainable development and poverty reduction. The Gateway offers users access to the most comprehensive database of development projects, an international procurement marketplace and knowledge sharing in key development topics. BA has agreed with the World Bank to collaborate

on this project in two phases. The first phase would be to gain ownership of the main Gateway site in Arabic and be responsible for Arabic translation. The second phase would be to design, develop and host the Egypt Country gateway in Arabic. Country Gateways are locally-owned and locally operated initiatives, which explore the use of information and communication technologies for development in their countries. By participating in and launching diverse online and offline initiatives, they enhance transparency, efficiency, dynamism and sustainability of the development process in their countries.

Virtual Visits

One of BA's objectives is to access the Library of Alexandria through virtual visits. ISIS has been involved in creating virtual visits to the entire BA complex, including the exterior and interior architecture in 3D. A prototype presenting the visiting tools functionality has been completed. The final version will include interactive views, a tour guide character, interactive objects, and 3D animations and narrations.

Another project is also taking place aiming at developing BA as a memento to the old Library of Alexandria, while utilizing contemporary Hi-Tech methods and capabilities to make it a World Resource Center. The project, funded by the Canadian International Development Agency (CIDA), will establish a Virtual Museum - *The Museion* - to present the Library as a heritage of humanity throughout its evolution. Interactive educational and promotional packages for students and visitors will be developed as well. Initially, a prototype will be implemented to include a *Virtual Exhibition of the History of Science* and a section for *Children's Library.*

A panoramic view of the Library taken from the pedestrian bridge

The Supercourse

www.bibalex.org/supercourse

The concept of the Supercourse is building a "metaschool" of a variety of themes that could train students through the Internet. Lectures are selected, published on the Internet and shared among scientists and teachers for personal use. A pilot project has been initiated featuring public health where 1000 of the "best" electronic lectures in epidemiology were published on CD and on the web. BA maintains a mirror site of the Supercourse (http: //www.bibalex.org/supercourse), thus ensuring high availability and reliability. By building a strong network of participating institutions and individuals, the target is to increase the current collection from 2000 to 100,000 lectures within one year and to one million lectures within three years.

Universal Networking Language (UNL)

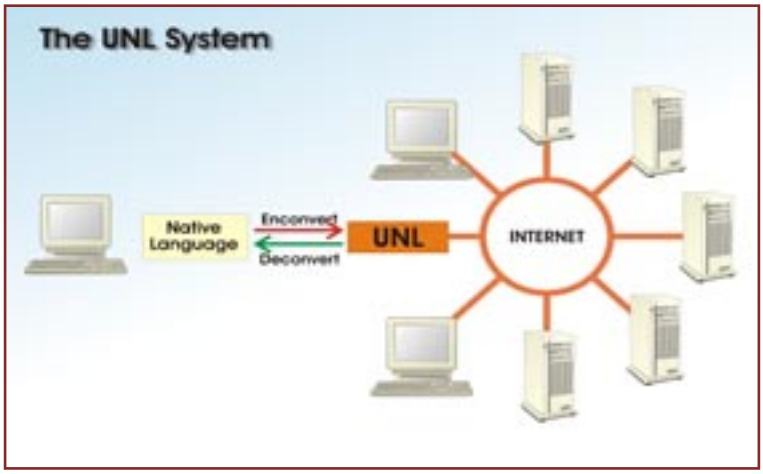
http://www.undl.org/unlsys/introduction.html

The mission of the UNL program, initiated within the United Nations and devised by the Universal Networking Digital Language (UNDL) Foundation, is to enable all people to generate information and have access to cultural knowledge in their native languages. The project is concerned with automatic translation based on the Interlingua concept for supporting multi languages. UNL is an artificial language attempting to replicate the functions of natural language in human communication. The core of a UNL system consists of a pair of software to bridge natural languages with UNL: an Enconverter, converting expressions from native language to UNL, and a Deconverter, converting expressions from UNL into native languages. UNL applications will vary including creating multilingual web pages, UNL encyclopedia, etc. Currently, 15 languages have been involved and a number of institutions have started to work on their respective native languages.

The UNL Editor application splash screen

In December 2003, BA hosted the International Conference on Convergence of Knowledge, Culture, Language and Information Technologies in cooperation with UNDL, the Royal Scientific Society of Jordan, the United Nations and UNESCO. In July 2004, partnership with UNDL was established and an agreement was signed in favor of BA to host the Ibrahim Shihata Arabic-UNL Center (ISAUC).

The Center is funded by the Arab Fund for Economic and Social Development (AFESD) in honor of the memory of Dr. Ibrahim Shihata. Dr. Shihata was one of the greatest international jurists and experts on international development of the twentieth century who had great contributions in promoting the Arabic culture. ISAUC will play the major role in designing and implementing the Arabic component in the development of this language and will act as an active language center for Arabic. A team has already been formed with members from BA and the best linguists from Egyptian universities and was trained by UNDL. By January 2005, three applications were designed and developed. The Knowledge Base (KB) Master Definitions allows linguists to create Master Definitions and Universal Words, which form the vocabulary of the UNL artificial language. The UNL Editor provides a user–friendly interface for the Enconverter and Deconverter applications. The UNL Graphical Annotation Tool provides a simple graphical user interface for linguists to manually tag sentence tokens. ■



The UNL system diagram