





INTERNATIONAL SCHOOL OF INFORMATION SCIENCE





INTERNATIONAL SCHOOL OF INFORMATION SCIENCE

Alexandria June 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	. 5
Projects	. 6
The Internet Archive	. 6
The Digital Laboratory	. 6
MILLION BOOK	. 7
DIGITAL ASSETS REPOSITORY (DAR)	. 8
Universal Digital Book Encoder (UDBE)	. 8
The Digital Library of the Modern History of Egypt	. 9
Gamal Abdel Nasser Collection	. 9
DESCRIPTION DE L'EGYPTE	10
Revival of L'Institut d'Egypte	11
Online Access to Consolidated Information on Serials (OACIS)	
for the Middle East	12
AL-HILAL DIGITAL COLLECTION	13
The Digital Library Federation (DLF)	14
My Book: Digital and Printed	14
La Main à la Pâte (LAMAP)	15
Hole in the Wall	16
Virtual Visits	17
The Supercourse	18
The Development Gateway of the World Bank	18
Universal Networking Language (UNL)	18
Arab InfoMall Database and Website	20

INTERNATIONAL SCHOOL OF INFORMATION SCIENCE

INTRODUCTION

The International School of Information Science (ISIS) research institute was founded in order to maximize creativity and foster innovations within the Bibliotheca Alexandrina (BA).

ISIS, guided by BA goals to:

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

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

The mission of ISIS is 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 the four corners of the world will gather to study and develop new technologies serving BA's mission to become 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 with staff either assigned from the Information and Communication Technology (ICT) Department or recruited 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 ISIS projects are outlined below.

THE INTERNET ARCHIVE

The Internet Archive is a complete snapshot of all web pages on every website since 1996 till today. In January 2002, the Internet Archive in San Francisco donated a copy to the Bibliotheca Alexandrina. The objective was to establish the second center worldwide holding a full copy of the Internet Archive. The archive mirror site maintained within BA has been fully operational and has been accessed with over 7.9 million hits during the past year. Synchronization with the Internet Archive in San Francisco has become faster and more reliable since the 155 Mbps bandwidth upgrade in January 2005. In addition, it is now possible to expand the archive collection by capturing web sites (crawling) from within BA.

An agreement to build the second generation machines for web archiving, the Petabox, has been finalized. The Petabox is a machine designed to safely store and process one petabyte (a million gigabytes) of data. New machines for the 2005 collection will be designed and manufactured locally.

THE DIGITAL LABORATORY

ISIS has built its own digital laboratory equipped with state-ofthe-art technologies offering specialized digital services required





http://archive.bibalex.org



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

The Internet Archive displaying an archived TV broadcast by the various projects. The laboratory is used for digitizing a wide range of media including slides in multi-formats, negatives, books, manuscripts, pictures, maps, audios and videos, 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 towards digitizing BA collections and

collections of other international libraries interested in pursuing the goal of "Universal Access to Human Knowledge". Many of the following projects are highly dependent on the Digital Laboratory to achieve their goals.

MILLION BOOK

Initiated by Carnegie Mellon University, the primary longterm objective is to transfer all books into digital format, in partnerships with other scanning centers worldwide. The project aims 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) are working together 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 provide content to ensure that the collection is extensive, diverse and multilingual. BA is taking the lead in scanning and digitizing 75,000 Arabic books in a three-year time frame. The digitization rate is gradually increasing to meet the target of scanning and processing 5000 pages/day/scanner (2 shifts), which will lead to approximately 25,000 books/ year. The project has also provided a test-bed supporting research on improved scanning techniques, optical character recognition (OCR), intelligent indexing, machine translation, and information retrieval.

A database for the books, metadata and digitization status has been designed and implemented, and standards have been set for the process of digitization in order to improve quality of the scanning, the processing and the OCR phases. The complete cycle of the workflow to produce digital books has been automated and integrated with the Library Information System. The database was further expanded to accommodate 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. During the past year, 12 OCR fonts were constructed and tested for accuracy, which exceeded 90% for 11 fonts. Additionally, 3 font groups are currently under construction.

Inside the Digital Lab



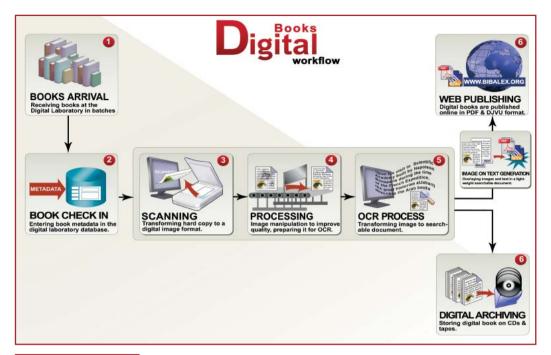
Cruse Scanner

Minolta PS7000 scanner for scanning face-up originals





7



Digitization Workflow

DIGITAL ASSETS REPOSITORY (DAR)

DAR was developed to create and maintain the digital library collections (whether already available in the Library or acquired from other institutions). The system acts as a homogeneous repository for all types of digital material and provides public access to digitized collections through web-based search and browsing facilities. DAR is also concerned with automation of the digitization workflow and its integration with the repository. The Digital Laboratory plays a major role in this respect.

The management tools developed within DAR will help the library preserve, manage and share its digital assets. The system is based on evolving standards for easy integration with web-based interoperable digital libraries.

UNIVERSAL DIGITAL BOOK ENCODER (UDBE)

For the purpose of electronic publishing of digitized material, UDBE was devised as a framework for the universal encoding of multilingual image-on-text documents, binding images and text in a compound format that allows retrieval systems to search the text layer and highlight hits on the original page images.

The UDBE framework renders it possible to utilize OCR results of any engine to compile image-on-text documents in any valid target format by adopting a proposed Common OCR Format (COF). The current implementation of the UDBE illustrates the concept by implementing an OCR Converter for Automatic Reader and Format Handlers for the DjVu and PDF target formats, making it possible to produce multilingual – namely, Latin, Arabic, and Persian – image-on-text documents in an automated fashion. The performance of the UDBE proved comparable to other systems capable of producing Latin imageon-text.

THE DIGITAL LIBRARY OF THE MODERN HISTORY OF EGYPT

This integrated digital library will include the collection of specialized libraries of eminent Egyptian politicians, authors and historians as well as content from all over the world related to the modern history of Egypt covering 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

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, 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;
- 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);

www.nasser.org

Home page of Gamal Abdel Nasser website

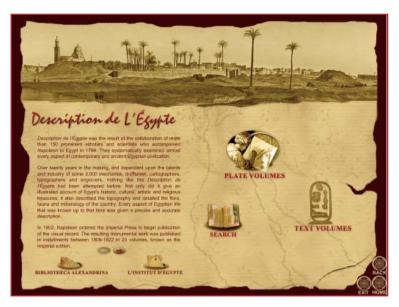


- The President's personal correspondence and other relevant documents in his own handwriting (596 documents under 145 topics);
- A complete archive of the special weekly column "Bisaraha" by the Egyptian writer and journalist Mr. Mohammed Hassanein Haikal documenting all major events during the period 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'Egypte 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



Description de l'Egypte, virtual browser main interface

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.

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 *L'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 two stages. The first project stage was completed in October 2004 where the application was provided on DVD in high resolution with French and English interfaces. The relation between the text and images were also established and rendered in a searchable form. In the second stage, the entire collection is to be available in an integrated searchable form on the web.

REVIVAL OF L'INSTITUT D'EGYPTE

BA is taking the initiative towards reviving *L'Institut d'Egypte* built in Cairo by Napoleon Bonaparte over 200 years ago. The *Institut* first scholars were responsible for 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

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





Institut d'Egypte built by Napoleon over 200 years ago

http://oacis.bibalex. org/ the Description de l'Egypte. Eventually, L'Institut d'Egypte became the focal point for scholarly 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.

L'Institut owns a huge collection (over 35,000 volumes) of rare and ancient references, books and periodicals in 5 languages (Arabic, French, English,

German and Russian). BA has suggested nine projects for its revival, among which is a project of digitizing its entire collection, thereby preserving it and making it available to the public. The efforts began 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 *L'Institut* will be digitized and made available to the public. This will be the first attempt to digitize and publish a collection of such rarity and value.

ONLINE ACCESS TO CONSOLIDATED INFORMATION ON SERIALS (OACIS) FOR THE MIDDLE EAST

Initiated by Yale University Library, OACIS creates 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 16 partners, 42 languages (with the top collections in Arabic, Persian and Turkish) and 23,000 unique title records.

A partner since August 2004, BA has been contributing to the project through:

- Enriching the OACIS catalog with BA catalog records and performing quarterly updates since January 2005;
- Digitizing serials and journals and implementing a contentretrieval web interface for the digitized material;
- Designing and implementing an Arabic web interface for the collection.

In January 2005, a mirror site of the system was launched at BA (*http://oacis.bibalex.org/*). Implementing a system for automatic uploading of BA records for updating the OACIS catalog is under plan.

AL-HILAL DIGITAL COLLECTION

The first issue of the collection was on 1 September 1892. Al-Helal is considered the oldest cultural journal in the Arab world, and the only regular journal that has been issued for more than a hundred years. It had a marked effect on the history of the Arab world in general and the history of Egypt in particular. It also played a leading role in modernizing Arab intellectual thinking, and opened new collaborations towards the cultural evolution.

ISIS is currently finalizing a significant project aiming to publish an exhaustive digital copy of the issues of al-Hilal since its first



Al-Hilal, virtual browser

publication in 1892, following an agreement concluded between the BA Manuscript Center and Dar Al-Hilal. The issues of each decade will be compiled on a CD and provided with necessary browsing and search tools.

THE DIGITAL LIBRARY FEDERATION (DLF)

On 20 May 2005, BA joined DLF as the first Strategic Partner from outside the USA and Europe, the first from the Middle East, and the second outside the USA (after the British Library). This will allow BA to participate in the world-wide effort of developing and promoting strategies and standards for creating and maintaining sustainable and scalable digital collections.

The Digital Library Federation (DLF), founded in 1995, is a partnership organization of academic libraries and related organizations that are pioneering the use of electronic information technologieas to extend their collections and services. Through its strategic partners, the DLF provides leadership for libraries by identifying standards and "best practices" for digital collections and network access; coordinating research and development in the libraries' use of technology; and incubating projects and services that libraries need but cannot develop individually. The DLF operates under the administration umbrella of the Council on Library and Information Resources (CLIR).

With 34 partners and 5 allies, the DLF identifies, documents, endorses, and promotes adoption of standards and best practices that support the effective acquisition, interchange, persistence, and assessment of digital library collections and services.

My Book: Digital and Printed

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.

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 metadata.

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 115 schools and 12 orphanages since November 2003.

http://archive.bibalex. org/mybook

Bookmobile unit at one of the schools





LA MAIN À LA PÂTE (LAMAP)

La Main à la Pâte is a french educational website which was initiated in 1996 by Georges Charpak, a 1992 Nobel prize winner for physics. It is managed by the French Academy of Sciences, and given the support of many partners to promote scientific investigation within the framework of primary school education. The website resources will help improve the teaching of science in primary schools by motivating teachers to place children in a position whereby they can experiment, observe, query and reason, opening them up to the beauty of the world around them and its intelligibility.

Recognizing the considerable role that sciences and technologies play in our lives, and the essential effect of this teaching in building our capacities, BA has partnered with the French Academy of Sciences to build an equivalent website of LAMAP in Arabic. The aim is to provide these especially tested resources for the Egyptian educational experience through a mirror website in the national language.

Recherche rebeldo Pranke dans tout le site TEDUJEL	Comparison de la contractiva de la contract						
Activités pour la classe Documentation Echanges Près de chez vous Actualités	Documents du jour						
	L'activité du jour La rose des vents Plus >>	Le projet Eratosti Mesurer I	1000	Demères questions Mardi 05 octobre 2004 Pourqui y at-11 des feuilles sur les arbres et pas des épines? Dimanche 03 octobre 2004 Comment traiter des énergies renouvelables Dimanche 03 octobre 2004 Quelle est la différence entre la châtaigne et le marron d'Inde ? Plus >>		Derniers messages Voir les archives de la list réseau-lamap Pl	
	Vous recherchez :						
	Une activité pour la classe		Une documentation		Des échanges		
	Acoustique Géologie Astronomie et espace Matière et matériaux Biologie humaine Mesures Ecologie Optique Electricité Technologie Energie Plus >> Evolution		Scientifique : Astronomie et espace Biologie Biologie humaine Ecologie Electricité Plus >:	Pédagogique : Cahier d'expériences Rôle du maître Plus >>	Des questions aux consultants Des listes de discussion		

My Book: Digital & Printed process cycle

LAMAP Website

BA will be contributing to the project through developing and launching the Arabic website with the same structure as the original, translating and adapting of material to Egyptian curriculum. It will also add back ends, new functionalities and administrative functions, build a network of Arabic scientists who will be contributing to the discussion forum, and promote the use of the Arabic site within the Egyptian community.

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.



HIWEL free Internet street kiosks

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.

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. Kiosks are monitored remotely to trace children activity.



This project will be accompanied by social and pedagogical studies analyzing the children's level and ability to explore computer technology, the interaction of the children with the project, their interests, and what they have learned from this project. The studies will also identify any changes in children's values, moralities and behavior that would result from them dealing with computer technologies; the effect of using computer technology in raising their culture, knowledge, and intellectuality; the most popular



computer applications, their usage amongst these children and how to develop them; and the impact of the children on their local economic activities after their exposure to computer technology.

The studies also aim to further identify how the project will interact with civil society organizations and other local organizations concerned with welfare of children. The study will include a survey of the target area and interviews with children and parents prior to and during project implementation, collecting and analyzing demographic, educational, and cultural data pertaining to the target community.

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. The virtual visit will be translated into Arabic and French, and the final product will be available on DVD.

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 A prototype of one kiosk has already been set inside BA

A panoramic view of the Library taken from the pedestrian bridge



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*.

THE SUPERCOURSE

The concept of the Supercourse is building a "metaschool" of a variety of themes to 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.

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.

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

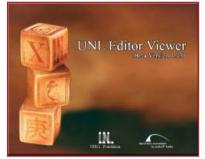
www.bibalex.org/

supercourse

UNIVERSAL NETWORKING LANGUAGE (UNL)

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.

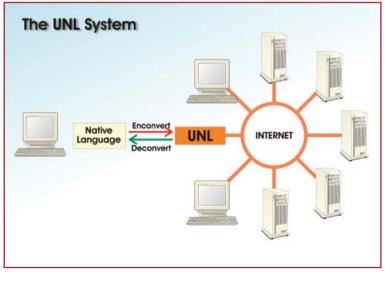
The core of a UNL system consists of a pair of software to bridge natural languages with UNL: an Enconverter, converting expressions from native language to UNL, and a Deconverter, converting expressions from UNL into native languages. UNL applications will vary including creating multilingual web pages, 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).

A team has already been formed with members from BA and the best linguists from Egyptian universities and was trained by UNDL. By June 2005, progress was achieved in: firstly, the Arabic Dictionary has been designed and the first version has been released with about 88,000 entries. Secondly, the workflow has been established for developing the Deco rules within the Arabic Deconversion component and Deco rules design has been completed. Finally, the design of the UNL Library Information System (LIS) is in progress to be ready for the World Summit on the Information Society (WSIS).



The UNL system diagram

ARAB INFOMALL DATABASE AND WEBSITE

The idea of this project is to emulate the role of a mall/center for non-governmental organizations (NGO's) and civil society organizations (CSOs) in a tri-lingual (Arabic, English, French) web portal under the name "Arab Info Mall" where each organization is presented as a kiosk/boutique inside the mall.

Launched in April 2005, the website beta version (*www. arabinfomall.org*) sheds light on NGOs/CSOs activities, successful projects and contact information, presented via a news bulletin, providing the opportunity for their local, regional and global exposure. Interested individuals have the opportunity to exchange opinions, contribute and share published reports through a discussion forum. Translation of the website into Arabic and French is in progress.



Arab InfoMall Website