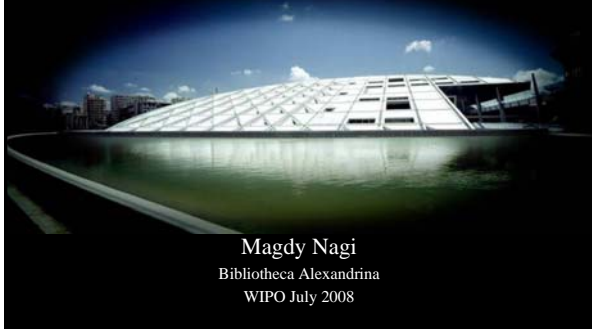


## International Workshop on Digital Preservation and Copyright



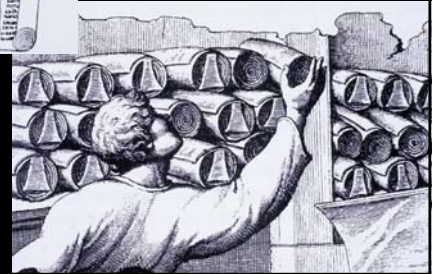
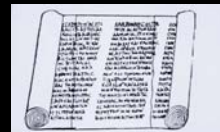
Magdy Nagi  
Bibliotheca Alexandrina  
WIPO July 2008



The  
library  
held about  
700,000  
scrolls,  
arranged  
in storage  
racks



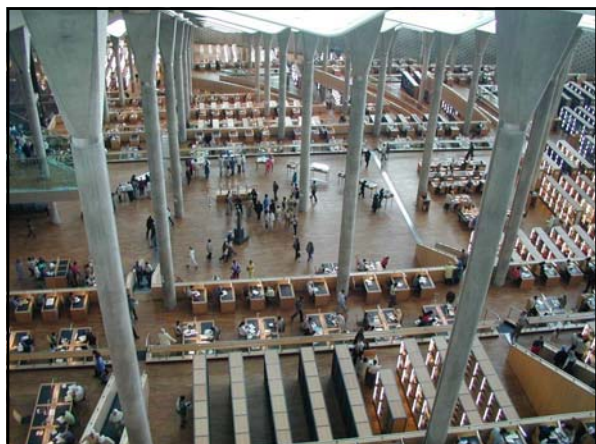
## Callimachus Pinakes



Little remains physically ...  
but  
it lives in the minds of all  
people...

## An Architectural Masterpiece





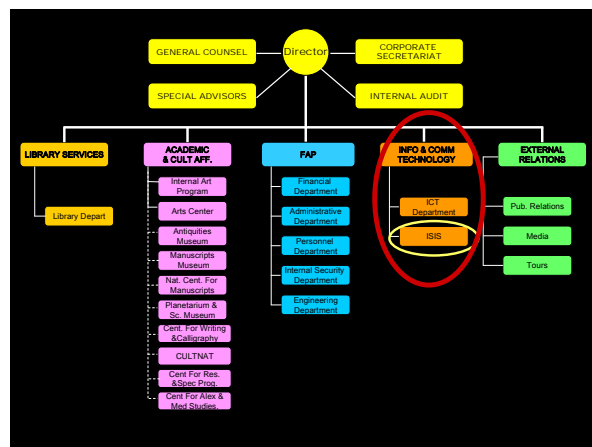
## Bibliotheca Alexandrina

- Window on Egypt for the world
- Window on the world for Egypt
- **A leading institution of the digital age**
- A vibrant center of intellectual debate – a space of freedom, for dialogue between individuals and civilizations



## ISIS Pioneering New IT Projects

All done in **partnerships** with powerful, imaginative proven partners



## Questions?

What is the difference between:

- Archiving
- Preservation
- Dissemination

The difference is minimal (or perhaps great!!), but they are linked together

**Committed to:**

Access to all information  
For all people  
At all times!

## Digital Object Archiving

Where a Copy / Copies can be kept and we are sure that:

- It will not change with time
- When the copy is damaged the archive keeper will be notified
- A new **exact** copy can be quickly made to replace the damaged one (e.g. a photocopy that is made from another photocopy is usually of lower quality)

The answer is: Put the **Master Copy** in digital format

## Digital Object Archiving (Cont.)

Digital copies can be kept on:

- CDs and DVDs:
  - Lifetime is less than 15 years
  - Can be damaged if not handled with care (scratches)
  - Long retrieval time
- Magnetic tapes:
  - Lifetime is less than 15 years
  - Reading the magnetic tape can damage the data
  - Very long retrieval time
- Online ☺  
Life time is > **2.4 Million** years according to SUN Systems

## Online Archiving

Archiving on **spinning disks**

- We can check as frequent as we like if part of all data is missing or damaged.
- We can use RAID technology to reconstruct the damaged data
- Might be costly but the cost of disks is becoming cheaper
- The archive is available 24x7 for retrieval and consultation

## Petabox

- Currently BA has ~ 4,000 TB
- Goals and current design points
  - Local computing to process the data
  - Multi-OS possible, Linux standard
  - Collocation friendly: requires our own rack to get 64/120/160 TB/rack
  - Shipping container friendly: Able to be run in a 20' by 8' by 8' shipping container
  - Easy Maintenance: one system administrator per petabyte



## Petabox

- Goals and current design points (cont'd):
  - Software to automate mirroring itself
  - Inexpensive design
  - Inexpensive storage



## Archiving projects at Bibliotheca Alexandrina

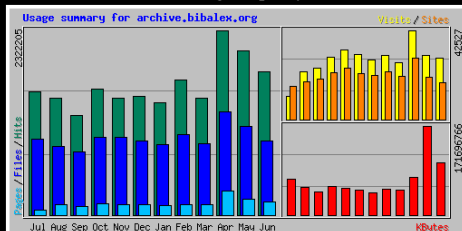
- Internet Archive
- Million Book Project
- Archiving of TV channels
- Modern History of Egypt
  - Description de l'Egypte
  - Nasser Digital Archive
  - Sadat Digital Archive
  - Mohamed Mahmoud Pasha
  - Botrosseya
  - Al-Hilal
  - L'Art Arabe
  - Suez Canal

## Internet Archive



## IA – Access Statistics

- The BA Archive site has been widely accessed reaching over 25 million hits during the past year



## IA – Progress

- The new web collection was indexed on the URL level and is now available to the public through the Wayback Machine: <http://archive.bibalex.org>



## IA – Progress

- An Automatic Cluster Synchronization (ACS) system was developed and was put to use in loading the Alexandria cluster with a selected subset of the web collection of 2007, a process that is currently in progress and has been making good utilization of the bandwidth

## IA – Future Work

- Continue deployment, configuration, and testing of locally-manufactured Petabox hardware
- Continue development of Automatic Cluster Synchronization (ACS) system and support of its operation to expand and maintain web collection data
- Explore special web crawling projects
- Explore non-web archiving projects
- Study new indexing structures

## IA – Future Work

- Invite researchers (computer scientists, linguists, etc.) to work on the data
- Build special collections that reflect the interests of Bibliotheca Alexandrina's patrons
- Develop software tools for parallel environments
- Continue work on enhancing cluster infrastructure

## Million Book Project

## Partners

- Carnegie Mellon University, US
- Internet Archive, US
- China, 20 centers
- India, 28 centers

More than 1,400,000 are now digitized

The new target is 10 M Books, but the project kept its original name

BA has digitized ~70,000 books

## Archiving of TV channels

In 2005, BA initiated digital archiving of TV channels

## Archiving took place on three steps:

- Installing required hardware.
- Video recording.
- Cutting and editing of recorded videos.
- Annotating the collected content

## Required hardware were installed



Satellite dish and receiver

## Satellite Video Card



Channels were recorded for 24 hours, seven days a week.



The project stopped, not because of the copyright, but because of difficulties in annotating the recorded content ☹

## Modern History of Egypt

- Description de l'Egypte
- Nasser Digital Archive
- Sadat Digital Archive
- Mohamed Mahmoud Pasha
- Botrosseya
- Al-Hilal
- L'Art Arab
- Suez Canal

## Modern History of Egypt

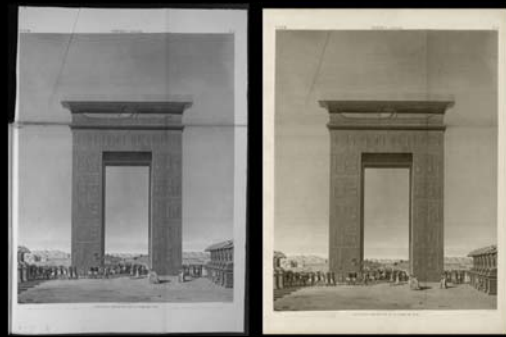
- Description de l'Egypte
- Nasser Digital Archive
- Sadat Digital Archive
- Mohamed Mahmoud Pasha
- Botrosseya
- Al-Hilal
- L'Art Arabe
- Suez Canal

## Digitization

The complete volumes of plates and text have been fully digitized.



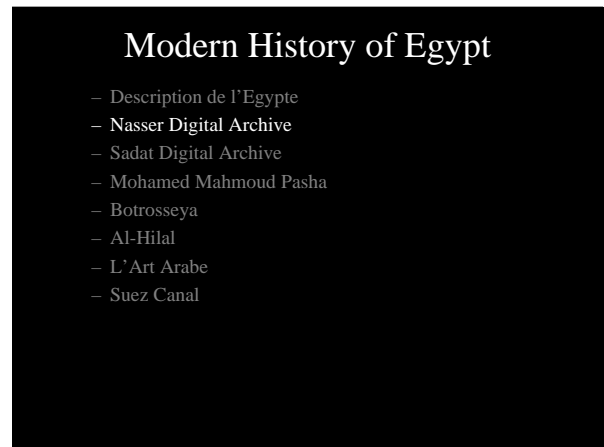
## Processing



## Virtual Browser-Plates Screen

- Plate Volumes
  - 11 volumes of plates
  - Recorded in three sections + Atlas
    - Antiquities
    - Modern State
    - Natural History
    - Atlas







## Nasser Archive – Objectives

- Digitize and publish the collection of Egyptian president Gamal Abdel Nasser



- Provide online access to his collection through a web based system mainly intended for research purposes and documentation

## Gamal Abdel Nasser Digital Archive



## Gamal Abdel Nasser Digital Archive

- Documents of Public Records Office, London, UK (53,000+ pages)
- Documents of the United State Department of State (30,000+ pages)
- 1,300+ speeches, audio and printed
- 51,000+ photos and 1,000 portraits
- 1,000+ videos (50+ hours)
- 1,200+ national songs
- 130+ Poems
- 140+ handwritten documents with 593 papers



## Gamal Abdel Nasser Digital Archive

- Articles published in the newspapers
- The decrees issued by the Revolutionary Command Council (RCC)
- Minutes of the Central Committee for Arab Socialist Union (ASU)
- The daily news of the President
- Archive of the "Bisara" articles by Mohammed Hassanein Haikal
- Caricature, stamps, coins and plastic arts illustrations
- Books written by and about Nasser

## Modern History of Egypt

- Description de l'Egypte
- Nasser Digital Archive
- Sadat Digital Archive
- Mohamed Mahmoud Pasha
- Botrosseya
- Al-Hilal
- L'Art Arabe
- Suez Canal

## El-Sadat Digital Archive

- Collection from
  - President Sadat's family
  - Newspaper agencies
- Collection includes
  - Pictures
  - Documents
  - Videos
- Workstations and scanners units are deployed at
  - Dar Akhbar Elyoum
  - Dar El-Mahfouzat
  - Dar Al-Helal



## Modern History of Egypt

- Description de l'Egypte
- Nasser Digital Archive
- Sadat Digital Archive
- Mohamed Mahmoud Pasha
- Botrosseya
- Al-Hilal
- L'Art Arabe
- Suez Canal

## Mohamed Mahmoud Pasha Digital Archive

- Digitizing the documents pertaining to Mohamed Mahmoud Pasha, one of the most famous Egyptian Prime Ministers
- Scope:
  - Digitize the entire collection of rich and rare historical documents and materials never been published before;
  - Provide it in searchable form for historians, politicians and researchers.
- Equipment have been installed in Cairo along with the software developed;
- Digitized 800 pictures
- Digitized 7,000 documents



## Modern History of Egypt

- Description de l'Egypte
- Nasser Digital Archive
- Sadat Digital Archive
- Mohamed Mahmoud Pasha
- Botrosseya
- Al-Hilal
- L'Art Arabe
- Suez Canal

## Botrosseya

- Digitizing the documents pertaining to the Botros Ghaly family
- The family has saved a large number of documents related to its political role since the late 1800's.
- Scope:
  - Digitize the entire multilingual (Arabic, English, French, German, Italian and Turkish) collection
    - 470 photos digitized
    - 15,000 documents
    - Collection of Botros Botros Ghali
  - Provide it in searchable form for historians, politicians and researchers.



## Modern History of Egypt

- Description de l'Egypte
- Nasser Digital Archive
- Sadat Digital Archive
- Mohamed Mahmoud Pasha
- Botrosseya
- Al-Hilal
- L'Art Arabe
- Suez Canal

## Al-Hilal

- The oldest continuously published cultural journal in the Arab world
- The only regular journal that has been issued for more than a 100 years
- It had a marked effect on the history of the Arab world in general and the history of Egypt in particular
- It played a leading role in modernizing Arab intellectual thinking, and opened new collaborations towards the cultural evolution
- Publish an exhaustive digital copy of the issues of Al-Hilal since its first publication in 1892
  - The volumes of the first 100 years are scanned, processed and indexed;
  - The issues of each decade are compiled on a CD including necessary browsing and searching tools.



## Modern History of Egypt

- Description de l'Égypte
- Nasser Digital Archive
- Sadat Digital Archive
- Mohamed Mahmoud Pasha
- Botrosseya
- Al-Hilal
- L'Art Arabe
- Suez Canal

## *L'Art Arabe*

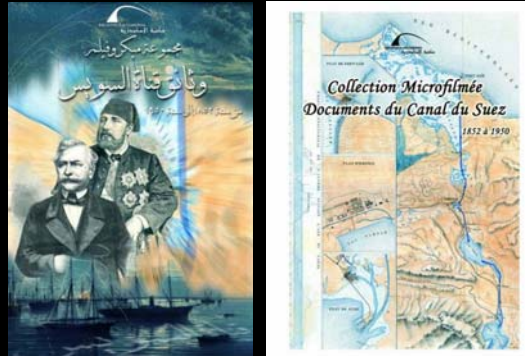
- *L'Art Arabe* is one of the most important books on the Islamic monuments of Egypt.
- The book is made up of four volumes compiled by the French Orientalist Prisse d'Avennes, one of the greatest pre-20th century Egyptologists.
- It was published in 1877, many of the examples produced in the book have since disappeared.
- The work postdates the *Description de l'Égypte* by 6 decades and can be seen in some ways as a commentary upon it.



## Modern History of Egypt

- Description de l'Egypte
- Nasser Digital Archive
- Sadat Digital Archive
- Mohamed Mahmoud Pasha
- Botrosseya
- Al-Hilal
- L'Art Arabe
- Suez Canal

## The Digital Memory of Suez Canal



## The Digital Memory of Suez Canal



## Digital Memory of Suez Canal

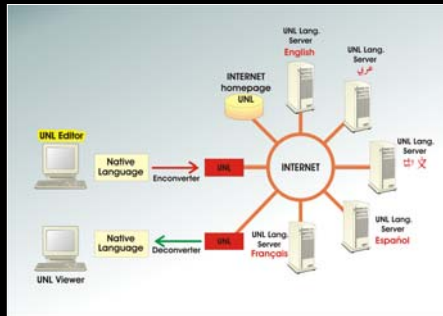
- November 2006: Friends of Ferdinand de Lesseps and Suez Canal Association donation to BA:
  - Documents in 2.5 million pages (equivalent to 1 kilometer) on 2,332 CD ROMs.
  - Two films of the construction and inauguration of Port Fouad.
- Scope:
  - Adding the digitized donations to the existing BA digital collection.
  - Digitizing the remaining items related to Suez Canal.
  - Publishing the digitized collection through a browsing application featuring searching and navigation tools.

## UNL

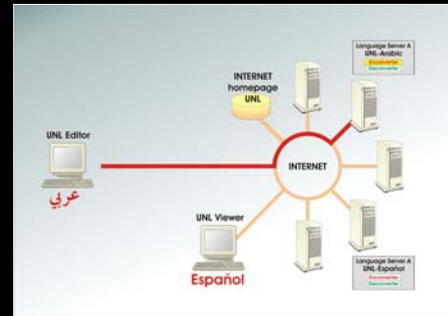
### Universal Networking Language



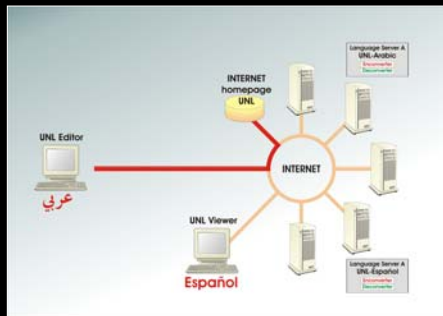
## UNL – System



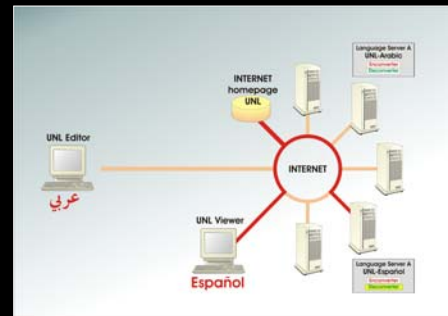
## Creating UNL Document



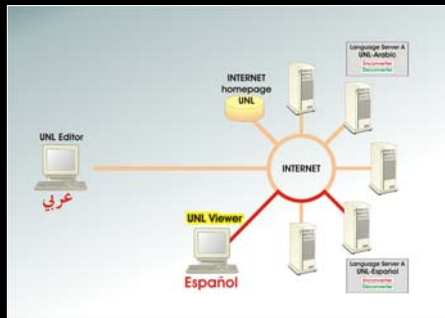
## Creating UNL Document



## Viewing UNL Document



## Viewing UNL Document



## UNL & Massive translations

- EOLSS is an Encyclopedia made of a collection of 20 encyclopedias, online, and in the form of e-books
- The number of pages of the encyclopedia is about 250,000
- A pilot project uses the UNL technology to (massive) translate 1000 pages only

Universal Access  
to All Knowledge  
for All People  
at All Times

Thank You