

Global BHL Quarterly Report

- Introduction:

The report documents the work done spanning March-May 2012, in terms of; Setting the infrastructure for the mirroring and replication of the original BHL book collection.

- Development of a Local-BHL system.
- Adding biodiversity material from the BA's collection which represents material from the Arab world in Arabic as well as other languages.

- Content Statistics:

- In addition to the two batches comprising 34,159 BHL books received by the BA on HDD, 70,566 books have been downloaded from the Internet Archive out of 105,359 books that are available on the Archive to date, thus yielding a current collection of **104,725** books.
- A complete version of the book files received either through the HDD or the Internet Archive is stored on the BA's Online Archive (OnA) where they are kept safely. OnA automatically mirrors the data and performs regular checks to make sure the data is intact.
- With 3,700 books currently in the pipeline, the collection of books published on <http://dar.bibalex.org> is expected to rise to almost 19,700 books (corresponding to **7,253,332 pages**) by the end of May 2012. The ICT department is currently working on a solution to accelerate the ingestion utilizing multiple parallel machines to increase the speed of the process from processing through to publishing.
- The BA is also working on adding featured biodiversity material from the Arab region in compliance to the BA's role in the EOL in Arabic project. A committee has selected an initial set of biodiversity material and is currently reviewing its copyright status and other BHL eligibility criteria. The collection spans content in Arabic as well as other languages.

- New software and functional development of a Local-BHL system:
 - Regarding the process of ingestion of books in BA's institutional Repository:
 - Several anomalies have been detected in the books downloaded from the IA: e.g. page images that were unprocessed (black edges, part of page, ...etc), pages that are out of order or not matching their XML coordinate files. Other alternatives have been investigated to rely on DJVU to extract images and XML files that are in order and in sync. Unfortunately, the images extracted from the DjVU suffered some loss of quality at high zoom levels. A Manual process could be needed to identify the books having such problems, and perform the necessary processing and publishing.
 - Regarding the development of BHL-Egypt Website
 - BA is working on the development of the BHL-Egypt website. This task involves user experience design, wireframes and functionalities to support the BHL community and tight integration with the Library's repository.
 - The BA team explored the differences between the metadata standards employed in the books downloaded from the IA and the books to be contributed by BA to identify sources of data for the website features. The investigation explored how different parties designate the volume information, relations between items, locations tags, ...etc.
 - The investigation also covered the BHL code on google code. An analysis of the database has been performed to complement the data extracted from the metadata files.
 - BA investigated Taxonfinder (FindIT and LinkIT), and is addressing how to integrate taxonomy functionalities into BA's book viewer.