

KOSMOPOLIS DIGITAL COLLECTION

Dimitrios Gavrilis, Panos Georgiou, Fiori Papadatou and John Tsakonas
Library & Information Service, University of Patras, Greece

INTRODUCTION

Kosmopolis is a digital collection containing the full-text content of twenty Greek periodicals from the mid-nineteenth century up to the beginning of the twentieth century. The project is being carried out by the Library & Information Service of the University of Patras within the framework of the TELEPHAESSA project (http://www.lis.upatras.gr/LIS/TELEPHAESSA/telephaessa_EL.shtml), in partnership with the Department of Greek Literature of the University of Patras and The Hellenic Literary and Historical Archive (ELIA, <http://www.elia.org.gr/>), Athens, Greece. The project is funded jointly by the European Social Fund and the Greek Ministry of Education within the Third Community Support Framework.

The tasks of the project include digitisation, bibliographical and scientific processing of the documents, and delivery to the users via a Web interface. The basic objectives of the project are:

- to extend public and scientific community access to rare and valuable content;
- the enrichment of digital content in the Greek language;
- to contribute to the preservation efforts of historical and valuable material;
- to create and promote new research challenges in Greek literature and history through the scientific processing of the digital content.

As a multi-purpose project, the Kosmopolis workflow and procedures were designed according to a step-by-step approach. Basic issues to be dealt with were:

- definition of the objectives and demands of the system, workflow design and resource management among the partners;
- definition of the basic and advanced standards and requirements for issues such as digitisation, technology and infrastructure, access to and storage of the digital content;
- development of the appropriate technical and organisational strategies for long-term digital preservation of the content;
- structure and functions of the appropriate Web interface;
- metadata schema;
- service evaluation models.

ORGANISATIONAL CONCEPTS

Partnership

Kosmopolis represents a collaboration between the following partners:
Library & Information Service (LIS), University of Patras (<http://www.lis.upatras.gr/>): LIS is the project coordinator with financial and administrative responsibilities. Furthermore, LIS played the major part in determining the technical specifications in digitisation and preservation issues, as well as in the design and development of the integrated digital collection system (e.g. the software and Web site). Finally, LIS stores the digital material in its storage facilities while it hosts, operates and maintains the appropriate equipment.

Department of Greek Literature, University of Patras: The department is responsible for the scientific coordination of the project (including title selection and indexing). It simulates the role of the potential end-user for design and test purposes.

Hellenic Literary and Historical Archive (ELIA, <http://www.elia.org.gr/>): This partner provides the original printed material and performs the initial digitisation phase.

Digitised material

The selected material belongs to ELIA's collection and includes periodicals of the nineteenth and early twentieth centuries. Basic selection criteria were (among others) the physical condition of the original, the availability and completeness of the series, and the absence of copyright problems.


Kosmopolis periodical & magazine titles

	Title	Coverage	Pages
1	Poikili Stoa (Ποικίλη Στοά)	1881-1914	7290
2	Euterpe (Ευτέρπη)	1847-1855	4610
3	Chrysallis (Χρυσάλλις)	1863-1866	2934
4	Ionios Anthologia (Ιόνιος Ανθολογία)	1834-1835	1218
5	Arothiki ton Ofelimon Gnoseon (Αποθήκη των ωφελίμων γνώσεων)	1837-1844	1565
6	Arothiki ton Ofelimon & Terpon Gnoseon (Αποθήκη των ωφελίμων & τερπνών γνώσεων)	1847-1849	316
7	Nea Zoi (Νέα Ζωή)	1904-1927	4000
8	Grammata (Γράμματα)	1911-1921	3040
9	Alexandriini Texni (Αλεξανδρινή Τέχνη)	1926-1930	2086
10	Argo (Αργώ)	1923-1927	726
11	Melissa (Μέλισσα)	1886-1887	392
12	Chloe (Χλόη)	1869	400
13	Eklekta Mithistorimata (Εκλεκτά μυθιστορήματα)	1884-1894	7200
14	Hmerologio Skokou (Ημερολόγιο Σκόκου)	1886-1918	15,000
15	Asmodeos (Ασμοδαίος)	1875-1880	1428
16	Estia (Εστία)	1876-1895	19,228
17	Deltion Estias (Δελτίον της Εστίας)	1877-1889	-
18	Ioniki Melissa (Ιωνική Μέλισσα)	1850-1852	388
19	Mi Xanese (Μη Χάνεσαι)	1880-1883	4574
20	Noumas (Νουμάς)	1903-1931	10720

Collection requirements

The collection was designed to meet some basic initial requirements:

- storage of digital content in appropriate file format to assure quality and conformity in relation to the original, compatibility with



current and emerging image processing technologies, and best storage space usage; use of proper storage devices for long-term storage conditions, easy copying and transferring and compatibility with current and upcoming hardware and software technologies; free and open access to digital content via the Web; content delivery in image, PDF and (in the future) HTML file formats; functional search, browsing and user help facilities

TECHNICAL SPECIFICATIONS

Digitisation

Taking into account the particular features and physical condition of the original material we used the following specifications to ensure good-quality digitisation:

The creation of two digital copies for each item (single page): a master file for preservation purposes and an access file for speedier delivery.

Master files were scanned at 300 dpi optical resolution, 1:1 size according to the original.

Master files were saved in jpeg format with the minimum possible compression.

Access files were created by processing master files to compressed jpeg files (25%) with the same resolution.

Image files are stored and delivered to LIS on DVDs.

Quality control was performed on the digitisation procedures to ensure that these standards were met.

Preservation of the digital material

For the long-term management and preservation of the collection and its content, LIS decided to adopt and develop gradually an OAI Reference Model according to ISO 1472 (Consultative Committee for Space Data Systems, 2002). In this initial project phase, LIS adopted a series of actions and procedures aimed at maintaining the best possible conditions for the digital content preservation. These procedures included the storage of master files on DVDs, while the compressed files for access were stored in a RAID disk storage system based on Fiber Channel 2-Gbit technology as well as on DVDs. We also maintained a stable storage room environment with air conditioning to maintain a mean temperature below 20°C. We had a specific plan for periodical checking of the physical condition of content and storage devices, and developed a strategy for the preservation of digital data, whereby data are refreshed (or migrated) onto new storage devices of the same or similar technology every three years.

Metadata schema

Dublin Core was the metadata schema chosen by LIS for the Kosmopolis collection for the following reasons: consistency with other LIS projects using the same schema, such as the subject gateway e-BGE (<http://www.lis.upatras.gr/ebge/>); the wider acceptance and support of DC as a standard metadata schema worldwide; the flexibility of the schema; it is accepted by and interacts with the OAI – PMH 2.0.

Kosmopolis digital collection management software

In the early stages of the project, Greenstone digital library software (<http://www.greenstone.org>) was selected for testing and for the pilot operation of the collection. Most of the appropriate modifications were carried out very successfully and a pilot version of the system was released in late 2003. Soon afterwards, problems arose in the areas of cataloguing, Web delivery and presentation and, most importantly, we realised that Greenstone could not implement the OAI-PMH 2.0. At that point the working group decided to proceed with the development of a new application that would meet all requirements and specifications.

Kosmopolis digital collection management software is a simple application that has been developed in PHP using a MySQL database. While it is running, it creates and stores dynamically the appropriate document images' navigation thumbnails under a cache catalogue. Access files are automatically resized according to the user's screen resolution (800x600 – 1024x768). Furthermore, users have the option of creating dynamically a PDF version of the document for saving and printing purposes. In this case, document images are resized automatically to fit onto A4 paper using the function `pdf_place_image` from the PDFlib library.⁴⁹

EVALUATION MODEL

Kosmopolis is the kind of digital collection that aims to support multiple education, research and cultural purposes. Therefore it has to be treated as an information service and we have to evaluate its use, its functionality, and finally (if possible) its effect on and outcome for the end-users. From an administrative point of view we have to define and/or forecast the real cost factors for developing, operating and maintaining the system as well as examine

⁴⁹ More information on processing PDF files on the fly can be found at <http://www.pdflib.com/>.

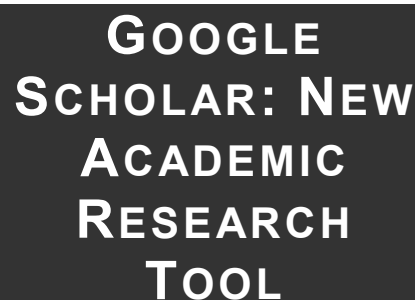
its viability. This approach demands an integrated evaluation framework focusing on these issues. In parallel, a detailed promotion plan has been designed aiming to reach all potential users worldwide.

In this initial phase we recorded, and are considering, a number of statistics and performance indicators such as number of accesses to collection, number of documents downloaded in total and per title, number of searches performed, downloads per various file formats (e.g. jpeg, pdf), among others.

CONCLUSIONS

The current version of Kosmopolis was launched in August 2004, with the content from three titles available to users via the Web. Since then the digital content of six more periodicals has been imported to the system, now totalling nine periodicals and about 7,500 documents. The rest of the content will be added gradually to the collection until the end of 2005. The promotion plan has already been set in action and by the end of 2006 all known potential users will be informed about the new service. The collection has been registered as an OAI-compliant Data Provider (<http://www.openarchives.org/Register/BrowseSites>) and it will gradually be registered with a number of known data harvesters, such as OAIster (<http://oaister.umdl.umich.edu/o/oaister/>).

Further reading and detailed information about the Kosmopolis project are available at: http://www.lis.upatras.gr/LIS/TELEPHAES/SA/dlib_EL.shtml, where the final report of the project can be accessed (Greek only).



GOOGLE
SCHOLAR: NEW
ACADEMIC
RESEARCH
TOOL

Adam Rusbridge, DCC
Development Analyst

The academic community produces a large quantity of literature that is of interest to fellow scholars and public alike. Without explicit direction towards a resource it can be difficult to discover additional and relevant academic information. The Google Scholar service (<http://scholar.google.com>),