Text & Data Mining for the National Library of Greece in consideration of GDPR

Dr. Marinos Papadopoulos
Attorney-at-Law

Mr. Konstantinos Vavousis, M.Sc.
IT Security expert

Mrs. Dimitra Hioti, M.A.
Information Scientist, Librarian

NLG & TDM in Greece

- Law 4452/2017 art.4(4)(b)
- Regulation 2016/679/EU (GDPR)
- Directive 2019/790/EU (Directive on Copyright in the DSM)
- NATIONAL LIBRARY OF GREECE (NLG) deployed TDM since FEB.2017
- Harvesting the *.GR*, *.EDU*, & *.COM* of the Web
- 18 TB of information

Working Stages Of Web Archiving In Greece By NLG

Stage I	Economic and technical study on the needs and content of the Greek web harvest. Study of international experience
Stage II	Definition of "Greek" sites to be mined
Stage III	Data Analysis of 1st web harvest to create a National Web Archiving System

1st web harvest:

broad crawl – national level: text data only

Stage IV	Installing and checking the operation of tools for all phases of national web archiving: extraction, archiving / classification and finally, user search and access): Heritrix for harvesting, Solr for indexing and Open Wayback for web site reconstitution. Netarchive Suite using.
Stage V	Developing a National Archiving System of Greek Web ("ΕΣΑΕΙ"): the Greek user/librarian interface

2nd web harvest:

broad - national level: text only)

thematic (text and images)

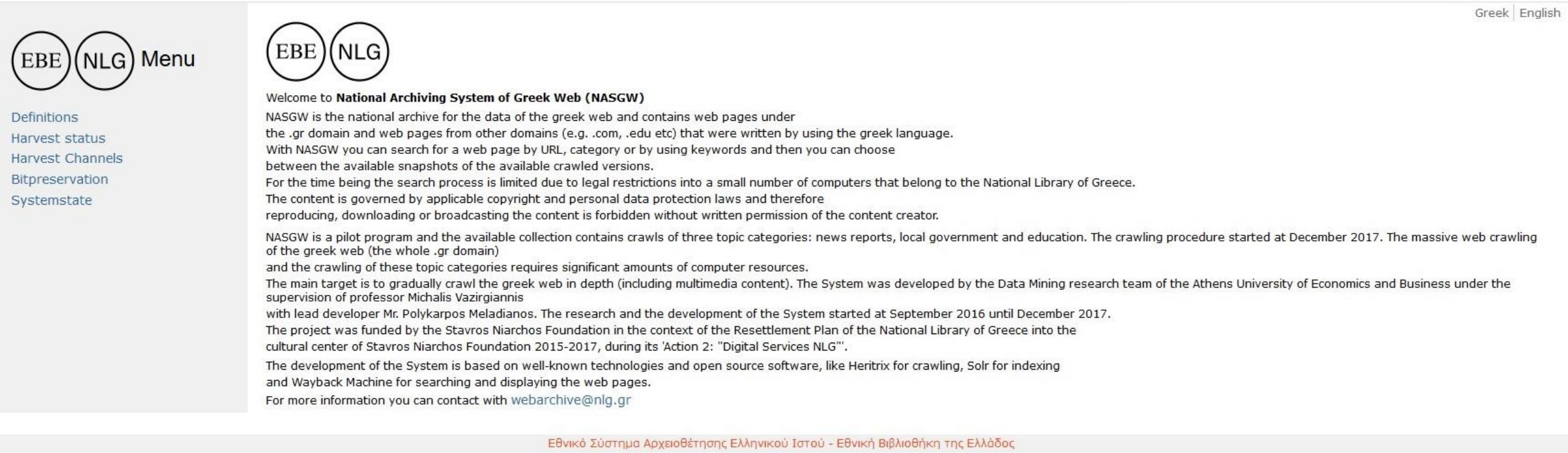
Harvesting

- Bulk harvesting of works in the Greek Web.
- Selective harvesting leveraging on criteria:
  1. Subject/Topic
  2. Creator/Provenance
  3. Type/Format
- The NLG Curator tool allows for either domain or selective harvests according to 6 predefined schedules.

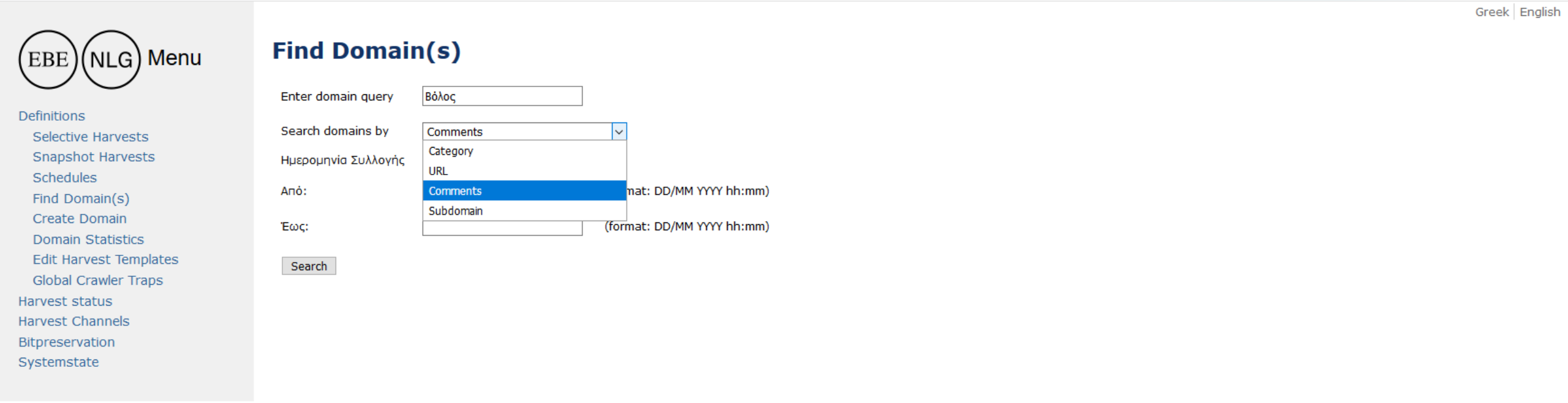
Preservation & Access

- Copies of harvested works are preserved in different servers in different buildings of NLG.
- Users can access works leveraging on the NLG Curator tool.

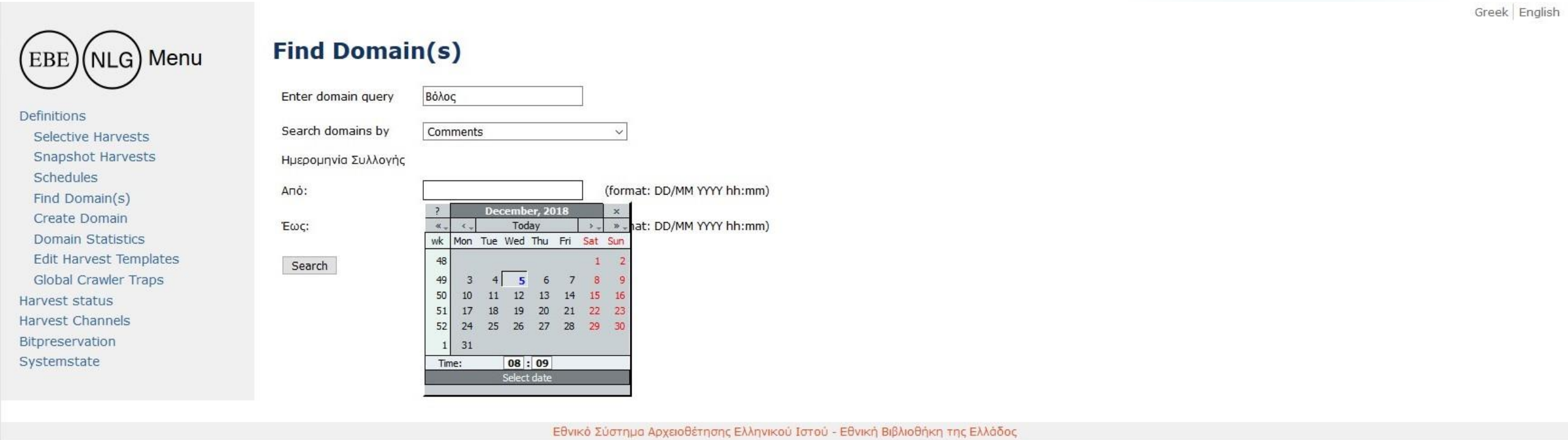
ΕΣΑΕΙ WEB ARCHIVE



ΕΣΑΕΙ WEB ARCHIVE



ΕΣΑΕΙ WEB ARCHIVE



GDPR vs SECURITY?

- Digital environments
- Strong Security Mechanisms

R.O.S.I.

- Minimize investments
- Heavy fines by GDPR

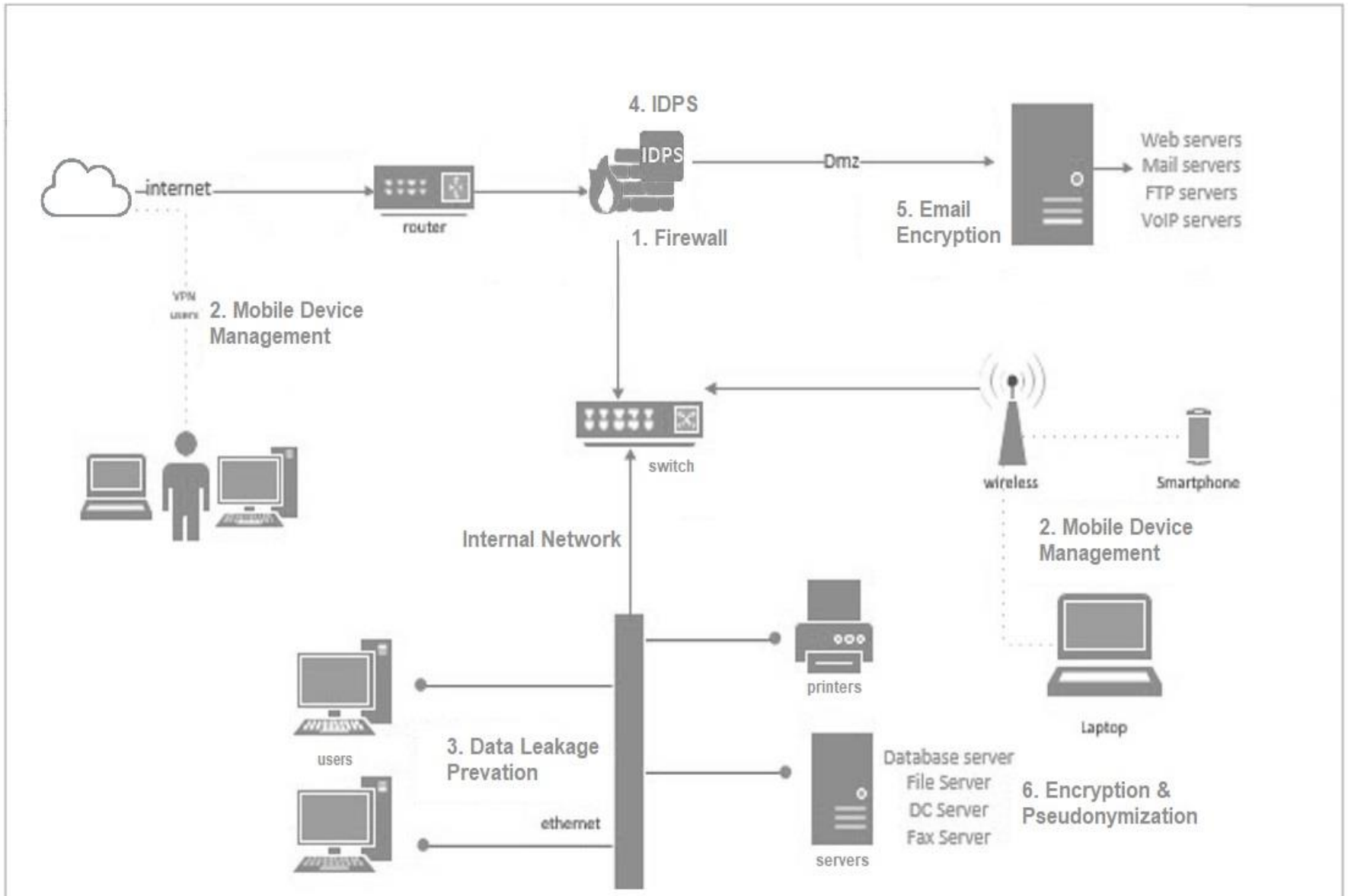
Strong Security Mechanisms

- Functions (Rec. 78, 108, 119)
- Processes (Rec.63, Art.4, 11, 28)
- Controls (Rec.37)
- Systems (Rec.49, 105, 151)
- Procedures (Rec.71, 88, Art.6, 12, 40, 41, 43, 47, 70)
- Policies (Rec.78, Art.4, 24, 39)

⇦ Personal and critical data protection

Optimal Infrastructure

- Firewall
- DLP
- MDM
- IDPS
- Email ENCRYPTION
- Database ENCRYPTION & PSEUDONYMIZATION



Authentication & Authorization

- Strong access control policies

Epilogue

- CYBER-SECURITY is top priority for GDPR

Text & Data Mining for the National Library of Greece in consideration of GDPR

contributors

Dr. Marinos Papadopoulos Attorney-at-Law	Mr. Konstantinos Vavousis, M.Sc. IT Security Expert Trust-IT Ltd.
Dr. Charalampos Bratsas Mathematician Aristotle University of Thessaloniki	Mrs. Eliza Makridou, M.Phil. Information Scientist, Librarian National Library of Greece
Dr. Michalis Gerolimos Information Scientist, Librarian National Library of Greece	Mrs. Dimitra Hioti, M.A. Information Scientist, Librarian National Library of Greece

Thank you