HEAL-Link activities and plans on annotating, organizing and linking academic content

Contributors:
N. Mitrou, N. Konstantinou, D. Kouis, P. Stavrou, D. Spanos

Presentation by
Prof. Nikolas Mitrou, NTUA
HEAL-Link chairman
HEAL-Link profile in numbers

Hellenic Academic Libraries-Link (known as HEAL-Link) was founded in 1998

- **Members & Associated Members**: > 50 Higher Education Institutes, Research Centers and other Public Sector Organizations
- **Cooperating organizations**: > 30 Public and Private Sector Libraries and Organizations
- **Academic & Research community**: > 13.000 academic staff (faculty members & researchers)
- **Students**: > 350.000 under/post graduate students
- **On line content**: > 14.000 scientific journals, 30.000 eBooks, various digital databases etc. > 4.000.000 bibliographic records
- **Current ongoing projects**: > 13 M€ funding for upgrading/creating digital services
HEAL-Link main action lines

- **Economise** resources
- Provide improved services for supporting teaching and research within academic units
- Set a **common strategy** regarding **access** to **digital material**, allowing information sources to be managed effectively state-wide
- Promote adoption of common performance standards and indices for library services
- Coordinate the development of the collections of the member-libraries
Expenditure for access to academic textbooks and scientific on-line sources

Not affordable anymore!
HEAL-Link's current activities towards Open Access

Two co-financed projects (national & EU funds):

- **Advanced Central Services of HEAL-Link's Open Access Digital Libraries** (2010-2015, 4.5M€)
- > 30 projects executed by Academic Libraries (HEAL-Link members) with a total budget exceeding 23 M€
- **Hellenic Academic E-books** (2012-2015, 8.3M€)
Advanced Central Services of HEAL-Link's Open Access Digital Libraries

**HEAL-Digital resources and Institutional Repositories service (HEAL-DIR)**

- Supporting and coordinating actions within the individual academic institutions in order to develop and/or upgrade their own institutional repository conforming to common standards which enable interoperability
- A common physical infrastructure and software platform for most Institutions
- A common set of metadata (healmeta)

- **HEAL Meta-search Service**
AMELib digital repository provides books to print-disable students through a set of tools for converting printed books to audio-books as well as other forms.
HEAL-Link Catalogues and Authorities/Indexing Service (HEAL-CAS)

- Electronic Authorities / Indexing Service
  - Greek Researchers and Research Institutes
  - Greek Scientific Journals
  - Greek Scientific Digital Resources
- ILSAS (Integrated Library System as a Service)
- Electronic plagiarism detection service
  - HEAL-HelpDesk
The service “Kallipos – Hellenic Academic Electronic books” aims to create and provide, in **open access**, a large number (more than 700 in the 1\textsuperscript{st} phase) academic textbooks as e-books.
Main features:

- Content mainly in the greek language
- Open access
- Multiple formats (pdf & Epub)
- Interactive & multimedia elements (video, sound, simulations etc.)
- Learning objects
- Complete metadata description (healmeta)
E-books vs e-courses

**An e-book**
- Is a linked collection of e-learning units, with a defined structure
- Is self-contained and portable within a standardized container, and presentable on special devices (e-readers)
- Has a unique theme and a defined sequence of reading
- Can contain multimedia elements and animations (**enhanced e-book**), as far as they can function locally
- Each stable edition can acquire a unique identifier (like the ISBN)
- Although outgoing links are allowed, all the basic building elements should be local, in order to assure the basic functionality in a non-connected mode

**An e-course**
- Is a loosely-bound set of e-learning materials
- Is provided usually by a Web server and presented by ordinary browsers
- Easy to incorporate multimedia elements, as well as animations in the form e.g. of **servlets** and **applets**
- Content can be changed dynamically. This may be a problem for assigning unique identifiers to individual content elements
- References to external components are easily handled
Epub: an e-book container

**Epub** is a distribution and interchange format standard for digital publications and documents:

- **Epub3** – the latest version
- Multiple resources wrapped into a single package
- *(XHTML)* HTML5 and SVG the primary content formats
- Multimedia and scripting supported (as defined in HTML5 and SVG)
- Human- and machine-readable navigation information
- Referencing document elements through the `epubcfi` (canonical fragment identifier)
- A minimal (but expandable) dc-based set of metadata
- Structural semantics
A typical Epub3 container structure

myBook.epub
- mimetype
- [META-INF]
  - container.xml
- [OEBPS]
  - package.opf
  - nav.xhtml
  - [css]
  - [fonts]
  - *.html
  - [svg]
  - [javascripts]
  - [images]
  - [mathml]
  - [audio]
  - [video]
  - [proprietary]

application/epub+zip

<?xml version="1.0" encoding="utf-8"?>
<container xmlns="urn:oasis:names:tc:opendocument:xmlns:container" version="1.0">
  <rootfiles>
    <rootfile full-path="OEBPS/package.opf" media-type="application/oebps-package+xml"/>
  </rootfiles>
</container>

<?xml version="1.0" encoding="UTF-8"?>
  <metadata xmlns:dc="http://purl.org/dc/elements/1.1/"
    dc:identifier id="pub-uid"
    dc:modified="2011-01-01T12:00:00Z"
    dc:language="en"
    dc:title="myBook-Title"
    dc:creator="N. Mitrou"/>
  <manifest>
    <item id="nav" media-type="application/xhtml+xml" href="nav.xhtml" properties="nav"/>
    <item href="css/main.css" media-type="text/css"/>
    <item href="chap01.xhtml" media-type="application/xhtml+xml" properties="scripted"/>
    <item href="images/im1.jpg" media-type="image/jpeg"/>
  </manifest>
  <spine>
    <itemref idref="chap01.xhtml"/>
    <temref idref="chap02.xhtml"/>
  </spine>
</package>
Ebub: Identifiers

- **Unique EPUB identifier**
  - Fully qualified URI (not required by the standard, but strongly recommended)
  - Example
    ```xml
    <package ... unique-identifier="pub-uid">
      <metadata xmlns:dc="http://purl.org/dc/elements/1.1/">
        <dc:identifier id="pub-id">urn:doi:10.1016/j.iheduc.2008.03.001</dc:identifier>
        <meta property="dcterms:modified">2013-11-26T12:00:00Z</meta>
        <meta refines="#pub-id" property="identifier-type" scheme="onix:codelist5">06</meta>
      </metadata>
    </package>
    ```
  - **Unique package identifier**: unique EPUB identifier + last modified date
    e.g. `urn:doi:10.1016/j.iheduc.2008.03.001@2013-11-26T12:00:00Z`

  - Defines a standardized method for referencing arbitrary content within an EPUB® Publication through the use of fragment identifiers
  - Syntax: `myBook.epub#epubcfi(path, [range])`
  - Example: `myBook.epub#epubcfi(/6/4[chap01ref]!/4[body01]/10[para05]/3:10)`
Why Epub?

Main assets

• Uses the web standards and technologies (HTML5, SVG)
• Is promoted by big names in the publishing industry
  (e.g. the Association of American Publishers/AAP)
Learning objects

**Definition:** “any entity, digital or non-digital, that may be used for learning, education or training” (IEEE)

**Features:**
- “small” entities (typical duration of presentation or reading: 5-20 minutes)
- self-contained in their own micro-container (file, folder, etc)
- uniquely identifiable, accessible and portable
- have an educational value
- reusable in compositions of larger LO (e.g. e-books)

**Examples:**
- interactive figures or maps
- algorithms
- mathematical theorems
- audio-video elements
- slide shows
- book chapters or sections
Learning objects (cont.)

- A novel implementation feature of the Hellenic Academic E-books
- Preserved within an e-book aggregator, with unique, dereferenceable URIs
- Self-contained, portable and re-usable for composing virtual e-books

- **Vocabularies-Metadata**: a combination of subsets from simple and widely used ones
  - **General (dc, ...)**
    - creator, title, creationDate, ...
  - **Structural (Epub, OAI-ORE)**
    - typeof {chapter, division, table, figure, toc}, partOf-hasPart...
  - **License (cc, ...)**
  - **Bibliographic (bibo, ...)**
  - **Scholar - learning (LOM, ...)**
    - prerequisites, interactivityLevel, difficulty, technicalRequirements, ...
    - hasSubject {subject headings}
Reusing LO for composing new e-books

When an LO is not self-contained but refers to or uses other local resources (style-sheets, javascripts, other contained LOs), care should be taken in order to take with it all the required resources, maintaining their relative position within the file structure.
**Issues to be taken care of**

- In putting resources from different containers into a shared space, there might be cases of name conflicts (same name, different functionality)

  **Possible solution:** Not use shared spaces. Isolate each LO along with its associated resources within its own micro-container.
An example from Wikipedia

Otto cycle
From Wikipedia, the free encyclopedia

See also: Otto engine and Four-stroke engine

An Otto cycle is an idealized thermodynamic cycle which functions as a typical spark ignition reciprocating piston thermodynamic cycle most commonly found in automobile engines. The Otto cycle is constructed of:

- Top and bottom of the loop: a pair of quasi-parallel processes
- Left and right sides of the loop: a pair of parallel isobaric processes

The book is a collection of relevant Wikipedia articles
The HEAL-Link e-book aggregator

e-books “kallipos”

Development steps
- (a),(b) Develop-populate registries of contributors (university staff) and Institutions
- (c) Compile subject-headings taxonomies and metadata schema (healmeta)
- (d) Archive learning objects and books, with the appropriate metadata
The HEAL-Link e-book aggregator (cont.)

**Learning Objects Repository**

- **Dspace platform**
  - Open-source software (initially by MIT and HP Labs → Dspace Foundation)
  - Content objects in a file system
  - Metadata in a relational database (Postgres)
  - OAI-PMH interface for metadata harvesting
  - Does not support the RDF model (Fedora does)

- **Same repository as for the rest of the academic content (publications, etc.)**
  [Institutional Repository]
Learning objects as Linked Data

**Guidelines**

- Each LO, being either within an epub container or stand alone, will be assigned a dereferencable URI
- The metadata (healmeta) will be mapped to elements of widely used vocabularies (dc, bibo, Dbpedia, ...) to the maximum possible extent. For the rest, new concepts/properties will be devised
- The thematic classification schemes will be organized by means of the SKOS vocabulary and interlinked with well-known taxonomies (LCSH, DDC)
- Outgoing and incoming links will be developed and advertized
- A live extractor may be developed for extracting information from the repository about modified or newly added LOs through the OAI-PMH interface (much like the DBpedia gets information from Wikipedia)
- ...
Learning objects as Linked Data (cont.)

**Incoming and outgoing links**

- Incoming links may refer to the Learning Objects preserved within the aggregator, like to any other content element on the web
- E-book readers, e-book composers or mashup applications will be able to discover, access and use the LOs by following these links
- Outgoing links may refer to any resource on the network
- In general, e-books are written and read as stand-alone content entities. E-readers do not require a network connection in order to present the content of an e-book. In such a connectionless mode of operation any outgoing links remain dormant; they can only be activated whenever the reader gets connected to the network
- The basic functionality and appearance of an e-book should not depend on external resources
Example

A Learning Object example
with links to external resources and navigation utilities

Nikolas Mitrou
Prof. NTUA
mitrou@cs.ntua.gr
Created: November 2013

Theme: Internal Combustion Engines and the Otto cycle

Abstract
In this example, a couple of illustrations and related information about a specific theme are presented, with the sole purpose of demonstrating useful features and presentation techniques when links to external resources are included for fetching content into a Learning Object dynamically. The content itself provided here about the theme is neither complete nor important for this demonstration.

• Presents two animated illustrations of the chosen theme
• Demonstrates the encapsulation of content from external resources (HTML pages or linked open data, e.g. DBpedia)
• Provides a set of navigation and comment compilation utilities
• Available by a server (as HTML) or within an epub3 container; so can be played by ordinary browsers or epub3-readers (e.g. Readium)

Institutional Repositories & library catalogs
Digital Library projects
Inference & Query engine
RDF store
Linked-data crawler/SPARQL engine
Well-established Ontologies (bibo, foaf, skos, dc, ...)
Ontology-thesauri constructors
New (heal) ontologies
Kallipos registries
People & learning objects
Policies
Institutions registries
Subject-headings taxonomy
Linked-data crawlers/2RDF convertors
extractors/2RDF convertors
healmeta, people & organization data, etc
Lists of University staff & Institutions
Lists of SH (GSRT)
SH (MARC21, SKOS)
SH (MARC21)
NLG SH (UNIMARC)
MARC2RDF
UNIMARC2MARC21
LCSH SH (MARC21, SKOS)
Kallipos
Healmeta, people & organization data, etc
Well-established Ontologies (bibo, foaf, skos, dc, ...)
Ontology-thesauri constructors
New (heal) ontologies
Kallipos registries
People & learning objects
Policies
Institutions registries
Subject-headings taxonomy
Linked-data crawlers/2RDF convertors
extractors/2RDF convertors
healmeta, people & organization data, etc
Lists of University staff & Institutions
Lists of SH (GSRT)
SH (MARC21, SKOS)
SH (MARC21)
NLG SH (UNIMARC)
MARC2RDF
UNIMARC2MARC21
LCSH SH (MARC21, SKOS)
Ideas for a Horizon 2020 project

?  

THANK YOU!