Emilia Kosa

Open access online resources and free web collections

/KVT-EN-03/

2019

EFOP-3.4.3-16-2016-00023
Az Óbudai Egyetem komplex intézményi fejlesztései a felsőfokú oktatás minőségének és hozzáférhetőségének együttes javítása érdekében

Table of Contents

1. Open Access (OA) Documents

Open Access (OA) means that the information is available for free of charge in the digital world. Open Access documents can be downloaded, copied, shared and printed by anyone with the source. The information is available fast and widespread through the worldwide web. Increasingly demands had shown that revolution was needed in the publication of research.

Why do the OA documents play a key role among the researchers and citizens?

- By open access documents the researchers can access the results easy and fast in the developed and developing countries.
- It ensures more visibility for researchers, programs, projects, results
- It increases the community of researchers
- The visibility of documents enhances the rate of citations
- The theoretical research can be adopted in practical version within a short time
- The research results are easy to approach for citizens and publicity
- The citizens receive values of their investment
- OA documents have role by the cultural mediation

The Open Access documents are available online for free of charge, but the highest journal publishers recognized that the change is needed in this section. They need to ensure to access their contents without subscription. This process was strengthened by the platforms what offer numerous publications from different editors. For scholarly it was important to archive and search their works. Due this processes the OA documents were formed different ways.

1.1 Types of Open Access Publishing

Green way: The authors or their institutions make self-archive about the publication. The authors offer it through their website or institutional repository.

Gold way: The journal publishers offer all articles in journal freely for everyone, therefore no need to subscribe to the journal. The publication fee as known article processing charges (APC) is charged for authors or their institutions.

Hybrid way: These journals contain open access and closed access articles as well for readers. The publication fee of open access articles are paid by the authors or their institutions and only receive free access to them.

Diamond/Platinum way: Those journal publishers which offer open access articles without APC and subscription fee referred to as Diamond or Platinum. These publishers create the foundation from external sources.

1. Process of Open Access Publishing

1.2 Article Processing Charges (APC)

In gold and hybrid open access publishing way the article processing charges (APC) are charged as publication fee. The readers can read, download these articles freely and the authors or mostly their institutions or grants cover these cost. APC is most common in medical and scientific journal and it depends on the impact factor of journals. It occurs lower percent in art and humanities journals.

The APC fee is widely range from under 100$ over 3000$, mostly include provision of online tools for editors and authors, article production and hosting, liaison with abstracting and indexing services, and customer services. If the authors have no supporting institution, the journal's publisher can offer various opportunities. Some of them can offer a waiver or discount or provide some assistance to find a suitable funding sources or agencies.

2. Digital Repository

Digital Repositories are a document-server to store, manage, preserve and provide access to digital content. The OAIS (Open Archival Information System) model provides an adequate foundation for designing standard-compliant repositories.

2.1 Open Archival Information System (OAIS) reference model

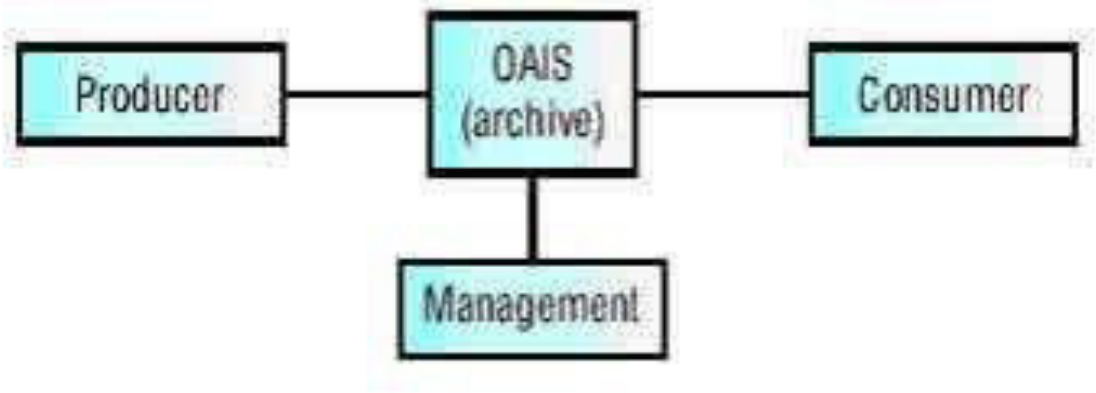
OAIS reference model is defined by the Consultative Committee for Space Data Systems (CCSDS) in the CCSDS Recommended Practice (2012) and the recommended standard is ISO 14721:2012. OAIS is an archive, consisting of an organization, which may be a part of larger organization, of people and systems that has accepted the responsibility to preserve information and make it available for a designated community.

OAIS reference model provides a conceptual framework for an archival system to preserving and access to digital information over the long term. The major purposes are:

- provides the concepts needed by non-archival organizations to be effective participants in the preservation process,
- provides a framework, including terminology and concepts, for describing and comparing architectures and operations of existing and future archives,
- provides a framework for describing and comparing different Long Term Preservation strategies and techniques,
- provides a basis for comparing the data models of digital information preserved by archives and for discussing how data models and the underlying information may change over time,
- provides a framework that may be expanded by other efforts to cover long term preservation of information that is not in digital form (e.g. physical media and physical samples),

- expands consensus on the elements and processes for long term digital information preservation and access, and promotes a larger market which vendors can support, and
- guides the identification and production of OAIS-related standards.

2.1.1 The OAIS Environment

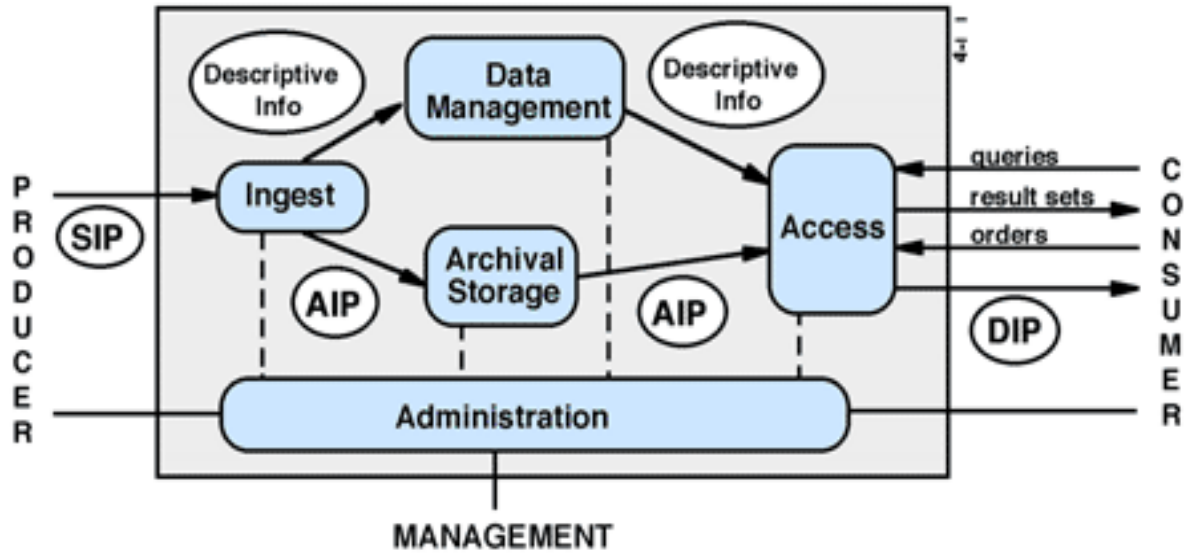


2. OAIS environment

The environment has 4 entities: producers, consumers, management and the archive itself. The producers supply the information that the archive preserves. The consumers use the preserved information. The special class of consumer is the designated community. Management is responsible for establishing the policies of the archive (e.g. determining types of information to be archived, identifying funding sources…etc.) It does not include the day-to-day administration of archive.

The model identifies 3 packages:

- SIP (Submission Information Package), which is sent from the information producers to the archive
- AIP (Archive Information Package), which is the information package stored by the archive
- DIP (Dissemination Information Package), which is the information package from the archive to the consumer as requested response.



3. OAIS functional entities

Ingest function is responsible for receiving the information from producers and preparing the contents for storage and management within the archive. It accepts the SIPs, performs the quality assurance on SIPs, generates an AIP from one or more SIPs and extracts Descriptive Information from the AIPs (metadata for search and retrieval…etc.) and finally the Ingest function transfers the AIPs to Archival Storage and the Descriptive Information to Data Management.

Data Management function coordinates the maintaining and accessing to Descriptive Information that identifies archive holdings and administrative data. This entity includes administering the archive database functions, performing database updates, performing queries on the data management data to generate results.

Archival Storage function is responsible for the storage, maintenance and retrieval of AIPs. This function receives the AIPs from Ingest and adding them to permanent storage, managing the storage hierarchy, refreshing the media on which archive holdings are stored, performing routine and special error checking, providing disaster recovery capabilities, and providing AIPs to Access to fulfill orders.

Access function supports Consumers to identify and obtain descriptions of relevant information in the archive, and delivers information from the archive to Consumers. This entity generates a DIP in response to a user request by the appropriate AIPs from Archival Storage, obtains the

relevant Descriptive Information from Data Management in response to a query and delivers the DIP or query results to Consumers.

Administration function manages the day-to-day operation of the archive system. This includes the negotiating submissions agreements with producers, performing system engineering, access control and customer service. It performs regular audits of SIPs to ensure that they meet archive standards. This function also serves as an interface between the archive and two components of OAIS environment: management and designated community.

2.2 Types of Digital repository

The repositories can classifiable in the following:

- Institutional repositories,
- Disciplinary repositories,
- National (government) repositories,
- Centralized (metadata) repositories.

Institutional repositories: The primary goal is to support the scholarly communication and provide open access to institutional research outputs, such as articles, thesis, dissertations, monographs, learning objects, conference proceedings. The institutional repositories are operated by an university or a research institution. The institution's scholarly research becomes global visibility. The unpublished and easily lost literature may store and preserve. For institutional repositories, three types of open-source software applications are most commonly used: DSpace, Eprints and Fedora.

DSpace was developed by MIT libraries and Hewlet Packard and was introduced in 2003. Eprints was developed by the computer science department of the University of Southampton, in the UK and it has been available since 2002. DSpace is widely adopted by libraries in the U.S.A. and the Eprints used mostly in the European and Australian library community. Fedora was named in 1997 at Cornell University the Flexible Extensible Digital Object Repository Architecture and use both in the U.S.A and Europe.

Disciplinary repositories: Disciplinary repository also known subject repository is an online archive containing works or data associated with these works of scholars in a particular subject area. Disciplinary repositories can accept work from scholars from any institutions. Disciplinary repositories shares the roles of collecting, disseminating, and archiving work with other repositories but is focused on a particular subject area. These repositories can include academic and research outputs. Disciplinary repositories usually cover one broad based discipline.

Featured disciplinary / subject repositories:

- arXiv: is an e-print service in the field of physics, mathematics, computer science, non-linear science, quantitative biology, quantitative finance and statistics
- PubMed Central: is a free full-text archive of biomedical and life science journal literature at the U.S. National Institutes of Health's National Library of Medicine.
- RePEc (Research Papers in Economics): is for research outputs in economics and related sciences.
- Cogprints: it is an e-print service, includes in any area of psychology, neuroscience and linguistics and many areas of computer science, philosophy and biology
- Humanities Commons: It is a trusted, non-profit network with peer-reviewed journal articles, dissertations and thesis, conference papers, abstracts. The network features an open-access repository, the CORE (Commons Open Repository Exchange).
- PeerJ Preprints: Preprints are not peer-reviewed. Preprints can be accepted if it fits the subject area, does not contravene any of their policies It publishes in these areas: biological sciences, environmental sciences, medical sciences, health sciences, computer sciences, chemistry.
- SocArXiv: It is an open archive of the social sciences, a free, non-profit platform for social scientists to upload working papers, preprints, and unpublished papers with the option to link data and code.
- ScienceOpen Preprints: It is a search and discovery platform. It offers a full suite of tools to peer-review and curate preprints. It allows the scientists set up a personal profile based on their ORCID.

National (government) repositories: Some National Repositories are financed by the country's government; others are industry-funded or use hybrid systems. These repositories are data banks to preserve and promote the country's natural resources data.

Centralized (metadata) repositories: A data dictionary refers as centralized repository of metadata. Metadata is data about data. It is a collection of tables with metadata.

The networked repositories can offer various features on user interface:

- the contents of repositories can search in a common platform as a virtual collection, regardless of the collection stored in physically
- users can set up a personal profile
- send automatically an electronic notice about the latest literatures
- prepare the publication list or bibliography of authors or research groups
- sort some versions of publications or link the stored data
- store citation and reference records

In the interest of avoid the screening all repositories one by one to achieve the required information, a protocol based on XML schema was developed by Open Archives Initiative.

2.3 Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH)

OAI-PMH is referred to as the OAI Protocol. The protocol is developed for harvesting metadata descriptions of records in an archive. In the late 1990s Herbert Van de Sompel (Ghent University) called a meeting to difficulties interoperability issues of e-print servers and digital repositories. Due to this meeting was developed an interface (Santa Fe Convention) that permitted e-prints servers to expose metadata so other repositories could identify and copy. The Open Archives Initiative was established by funding of Coalition for Networked Information and Digital Library Federation. The OAI-PMH version 1.0 was introduced in 2001.

OAI-PMH is based on client server architecture, in which "harvesters" request information on updated records from "repositories". Requests for data can be based on a datestamp range, and

can be restricted to named sets defined by the provider. Data providers are required to provide XML metadata in Dublin Core format, and may also provide it in other XML formats.

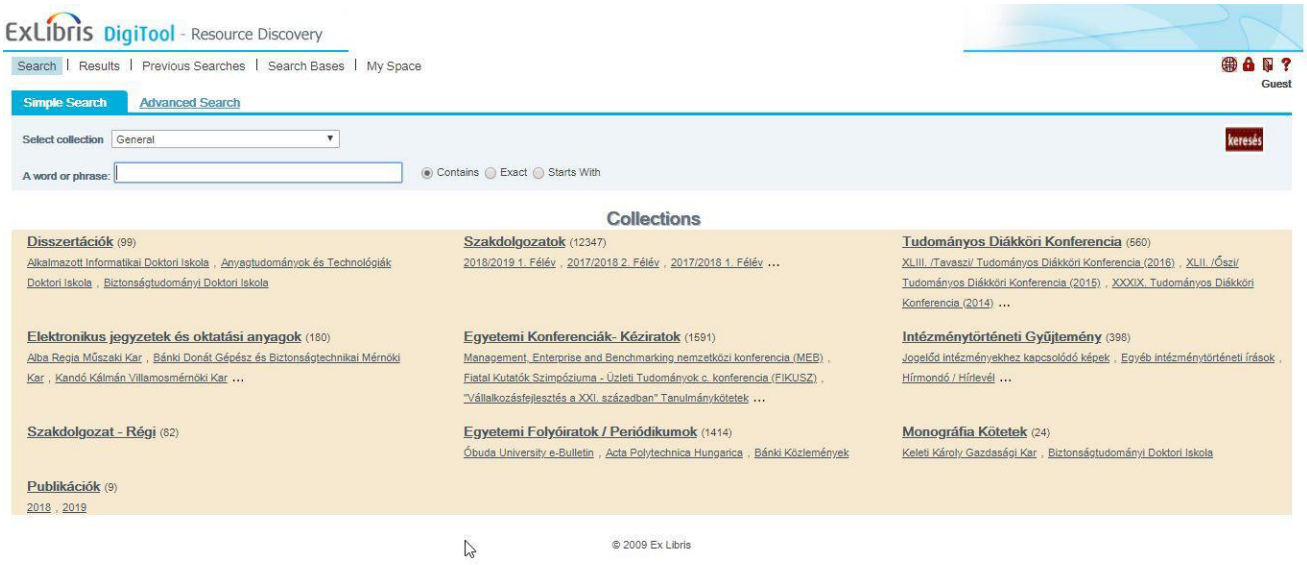
Numerous archives and software systems support the OAI protocol including arXiv, CERN document server, Fedora, Eprints, DSpace.

3. Description some institutional repositories

Institutional repository supports the scholarly, research and teaching work in the universities, research institutes. Academic and research outputs are available among the members in the community of institutes and for the public as well. The institutes by the Common Creative License may decide about the visibility of outputs. Some works can be visibility for the public and others only among the members of institute.

3.1 Digital Repository of Óbuda University (ÓDA)

ÓDA is full-text database that preserving, representing and providing the scholarly results of University. ÓDA contains more than 18000 documents including bachelor and master theses. The collection includes more Ph.D. theses, theses for Scientific Students' Associations, conference papers and academic publications. An institutional history collection is also part of the repository, in that the documents of Óbuda University and of legal predecessors are available and researchable. Due to historical documents are available mostly in printed form, the archival collection can be enhanced by the digitalization project of University Library.



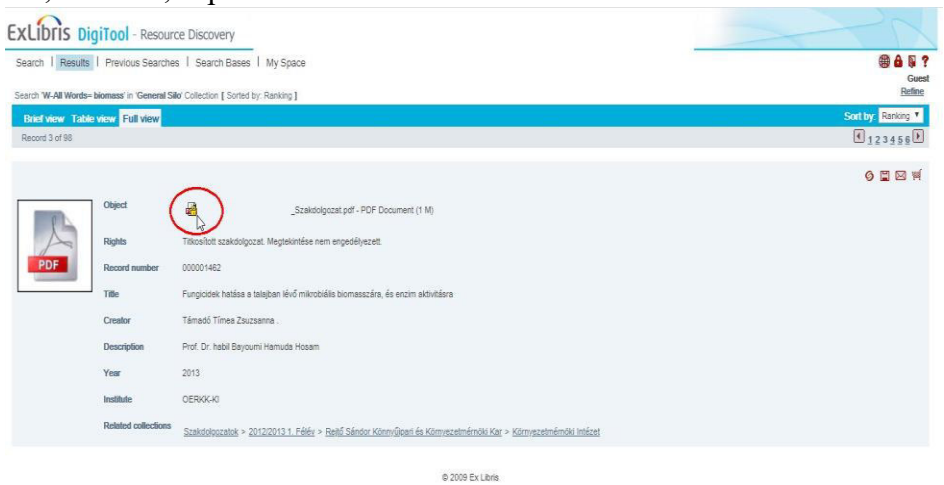
4. Digital Repository of Óbuda University

The University applies Digitool software by ExLibris that is based on Eprints. Digitool can use all types of digital documents such as text, image, video and audio documents. The relevant metadata can be linked together therefore hierarchical structure can be created. A metadata record can be linked more digital object. The software offers both metadata and full-text search. The search result can be sorted and ranked and the digital object appeared by standard browser or an application. The software supports several protocols and technological forms (e.g. MARC21, Dublin Core, UTF-8, OAI-PMH, z39.50..etc.)

ÓDA includes the following collections:

- Dissertations
- Electronic teaching materials
- Bachelor and master theses
- Conference paper
- Journals are published by University
- Theses for Scientific Students' Association
- Institute Historical Collection
- Monographs

By the search result we can set up the view (Brief view, table view, full view). The bachelor and master theses can read in full-text format via University network, from 'outside' can search metadata only. In that case the thesis is encrypted there is no accessible. The record contains title, author(s), year, institute, supervisor's name.

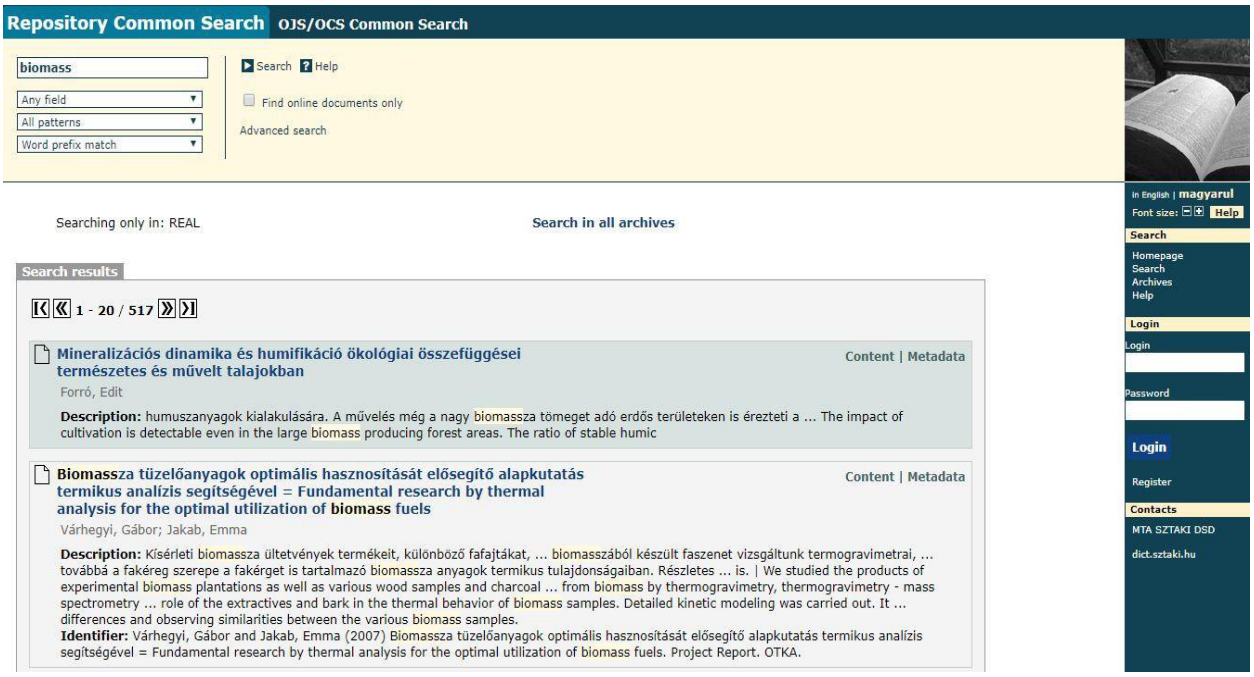


5. Encrypted thesis in ÓDA (small icon is unlocked)

3.2 REAL Repository

The Real Repository is operated by the Library and Information Centre of Hungarian Academy of Sciences. It includes full-text content, supports the OAI-PMH protocol and based on Eprints Sotware. The repository offers the following collections:

- REAL: for research output of projects supported by grants from National Research, Development and Innovation Office
- REAL-D: contains dissertations of CSc and DSc of the Hungarian Academy of Sciences
- REAL-PHD: contains dissertations of PhD at Hungarian universities
- REAL-EOD: contains books are not covered by copyright law
- REAL-J: contains full text periodicals published by the Hungarian Academy of Sciences
- REAL-MS: contains full-text manuscripts from the Oriental Collection
- REAL-R: contains full-text books from the Oriental Collection

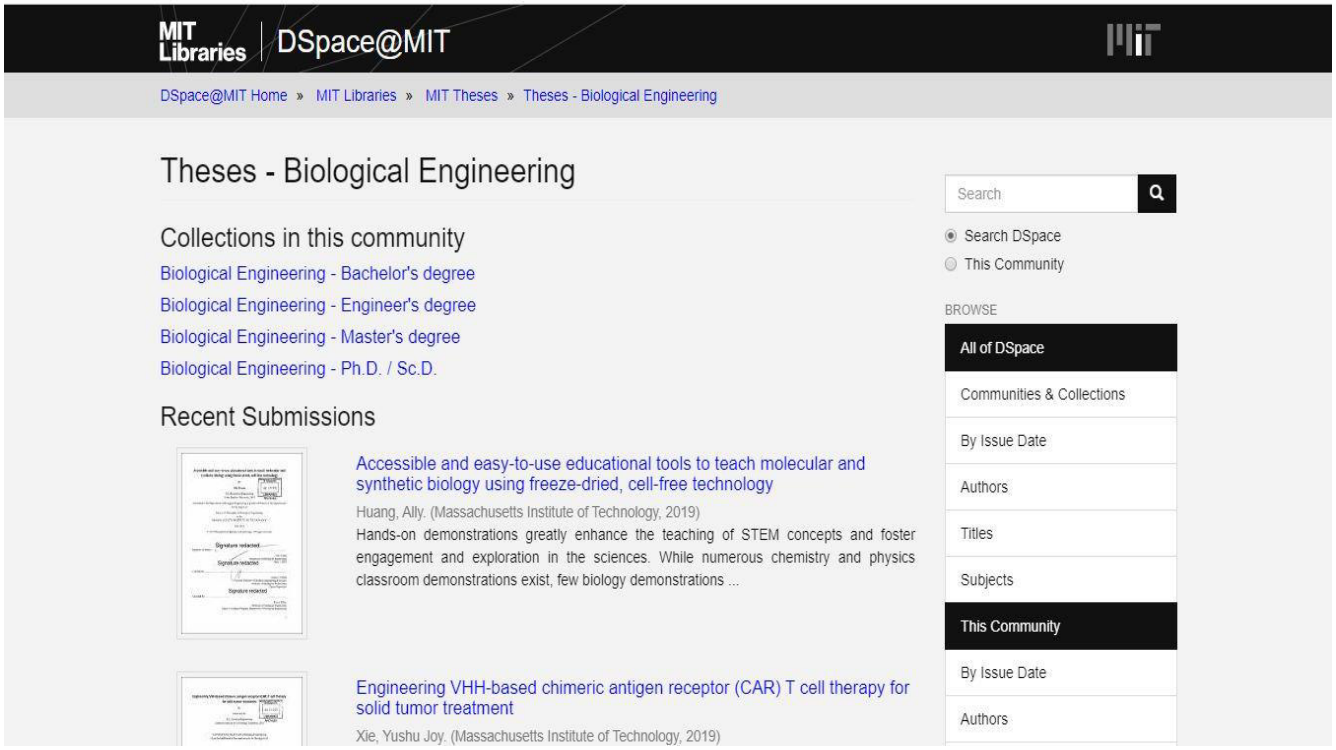


6. Sample search for 'biomass' term in REAL Repository

3.3 DSpace@MIT

DSpace@MIT is a digital repository for MIT's research, including peer-reviewed articles, technical reports, working papers, theses, conference papers, images, preprints and research datasets. It is a service of MIT Libraries to provide MIT faculty, researchers a long-term storage for their digital research and teaching output. The collection is more than 90000 works.

The collection of MIT Thesis in DSpace@MIT contains more than 44000 selected theses and dissertations from all MIT departments. The theses are openly accessible to anyone. These are print and copy, we can order a softbound copy. Theses have been scanned by Document Services of MIT Libraries or submitted in electronic format by thesis authors.

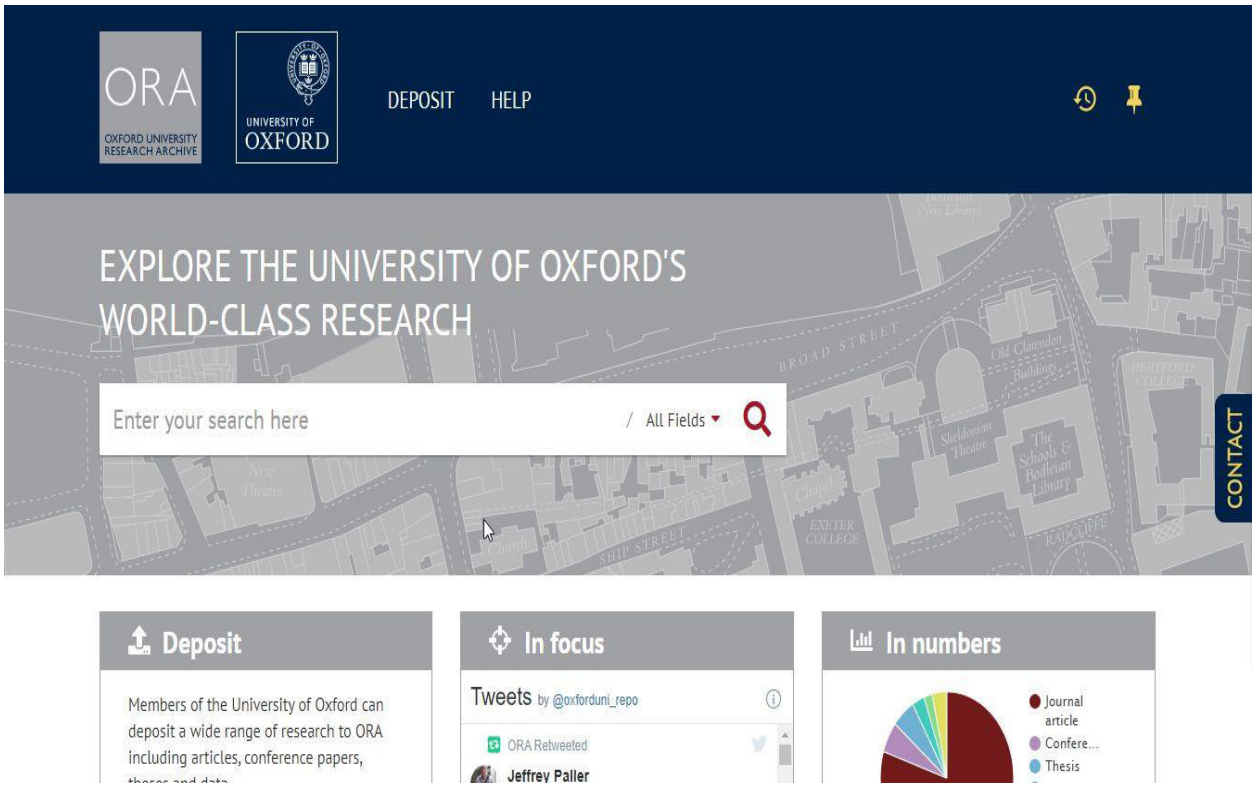


7. MIT Thesis collection in DSpace@MIT

3.4 Oxford University Research Archive (ORA)

The Bodleian Libraries' Oxford University Research Archive (ORA) is a repository of research materials produced by members of the University of Oxford. ORA was established in 2007. The technical platform for ORA is Fedora, a modular open-source repository system. The metadata is from other subject-based repositories such as PubMed Central, arXiv, Scopus and other publication within University. A dedicated team at the Bodleian Libraries supports the addition of items to the repository by members of the university, verifying metadata and license conditions. Ora provides three ways to deposit the research outputs of members:

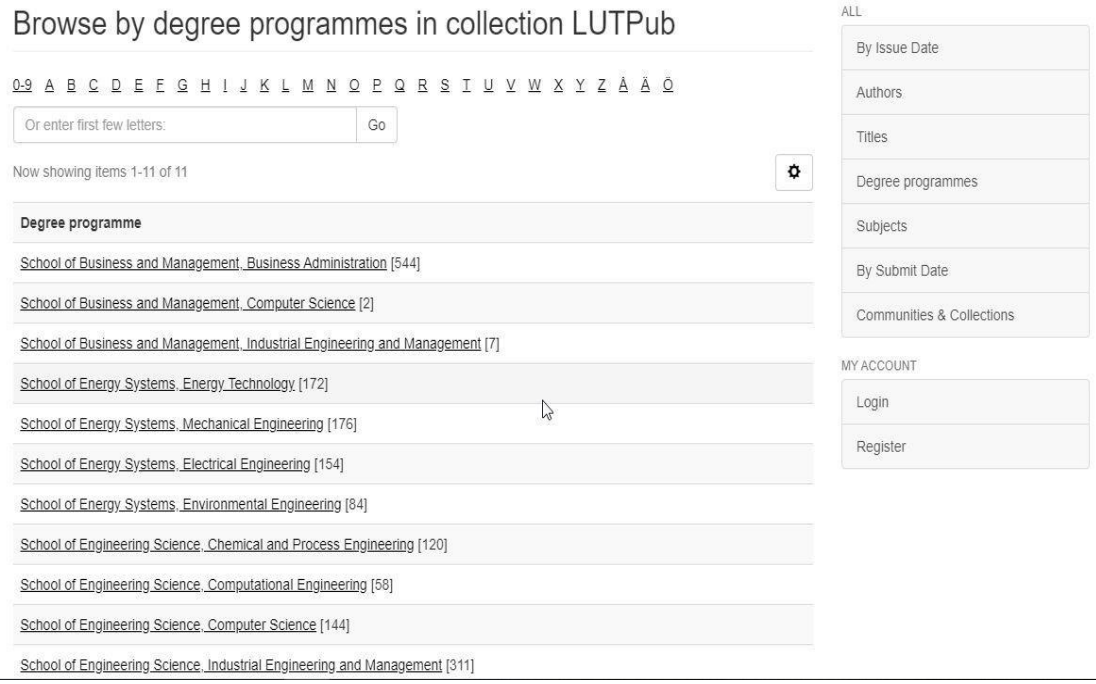
- Deposit research: for conference papers, journal articles, book chapter, reports…etc.
- Deposit theses: for research theses
- Deposit data: any type of digital research data



8. Homepage of ORA

3.5 LUTPub

LUTPub is the publication repository of LUT University (Lappeenranta-Lahti University of Technology) LUTPub contains Bachelor's and Master's theses, doctoral dissertations, research datasets, teaching materials and other research publications and outputs. Researchers can publish their articles in open access in LUTPub. There are browse options by author, title, issue date, degree program, subject, submit date and communities & collections.



9. Browsing in LUTPub

4. Open Access Resources

The Open Access has provided a hand in the spread of repositories, because transforms the scholarly communication. International special open access databases regardless of discipline appeared. The contents of Open Access Databases are under Common Creative license.

4.1 Directory of Open Access Books (DOAB)

Directory of Open Access is a service of DOAB Foundation. It is a non-profit, legal entity under the Dutch law, established by OAPEN Foundation and Open Edition. The Foundation is based at National Library in The Hague. By establish the OAPEN Foundation developed it close cooperation with Lars Bjørnshauge and Salam Baker Shanawa (director of SemperTool). SemperTool is responsible for Directory of Open Access Journal (DOAJ). OAPEN Library is an international initiative dedicated to Open Access monograph publishing.

The main purposes of DOAB are to enhance discoverability of Open Access books. "Academic publishers are invited to provide metadata of their Open Access books to DOAB. Metadata will be harvestable in order to maximize dissemination, visibility and impact. Aggregators can integrate the records in their commercial services and libraries can integrate the directory into their online catalogues, helping scholars and students to discover the books."[1] The DOAB is a great chance for academic publishers, who publish in open access to become more visible and accessible.

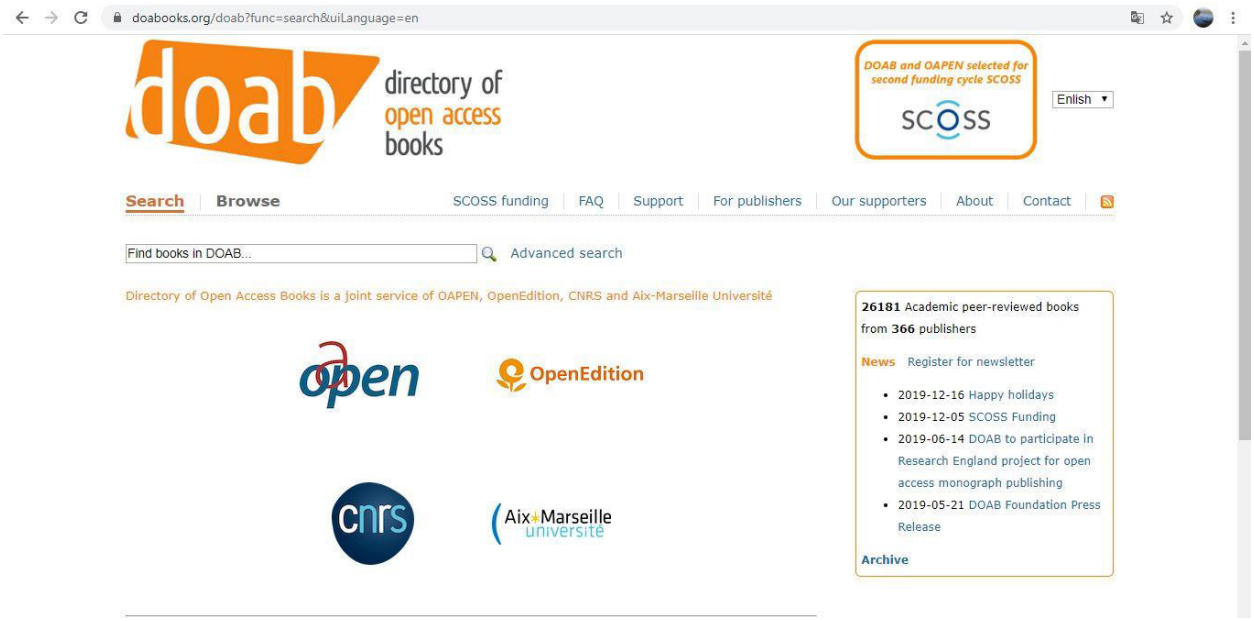
DOAB supports OAI protocol for metadata harvesting, therefore the librarians and service providers can harvest the metadata from DOAB to their records, collections and catalogues.

[1] https://www.doabooks.org/doab?func=about&uiLanguage=en

If a publisher would take part of DOAB, they should keep on eye some requirements. Two main requirements are:

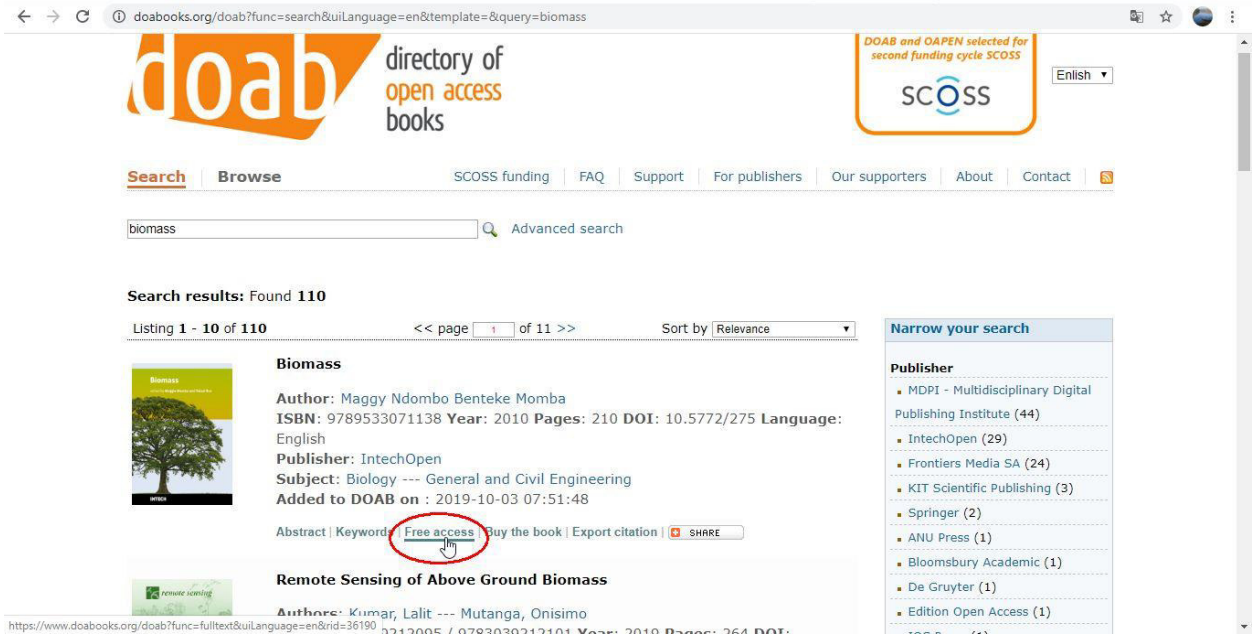
- The academic books in DOAB shall be available under an Open Access license and Creative Commons license as well.
- The academic books in DOAB shall be peer reviewed prior to publish

The publisher should fill up an application form and after can upload the book(s).



10. DOAB homepage

The database offers both search (simple and advanced) and browse. The search results can be sorted by relevance, year and date added. By records can be founded some data: title, author, ISBN, published year, pages, DOI, language, publisher, subject and the date when added to DOAB. By every record we have option to export citation to Refworks, EndNote or Reference Manager. Free access is linked directly to the record.



11. Search results in DOAB

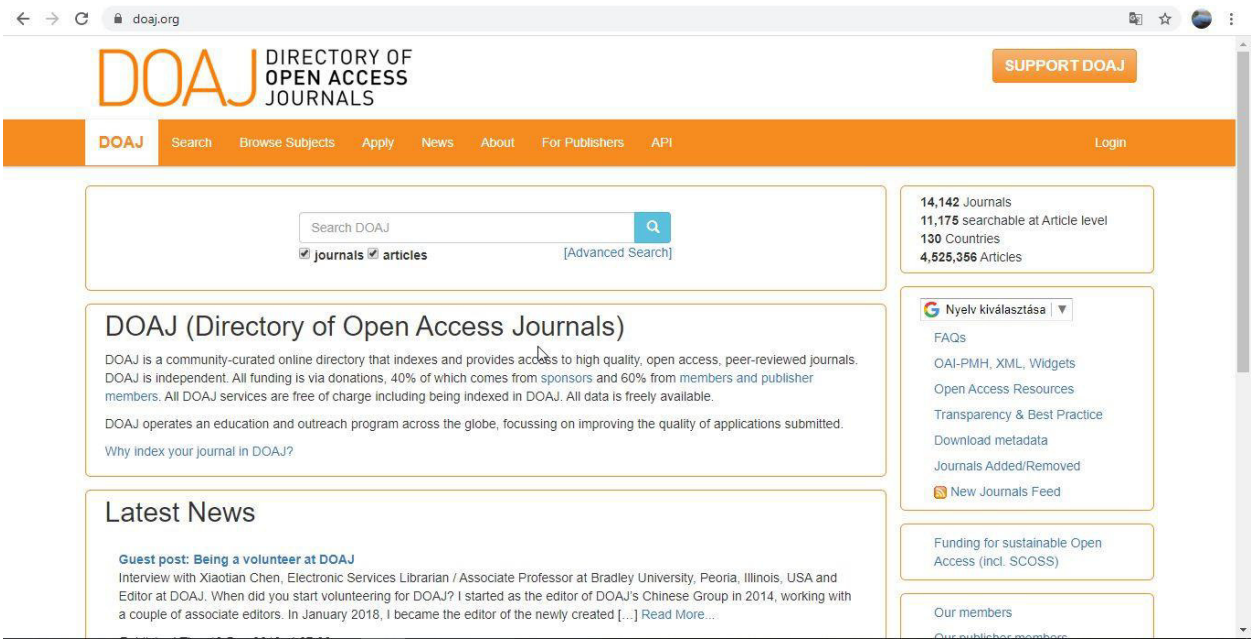
4.2 Directory of Open Access Journals (DOAJ)

DOAJ's mission:

"DOAJ's mission is to increase the visibility, accessibility, reputation, usage and impact of quality, peer-reviewed, open access scholarly research journals globally, regardless of discipline, geography or language. DOAJ will work with editors, publishers and journal owners to help them understand the value of best practice publishing and standards and apply those to their own operations. DOAJ is committed to being 100% independent and maintaining all of its services and metadata as free to use or reuse for everyone."[2]

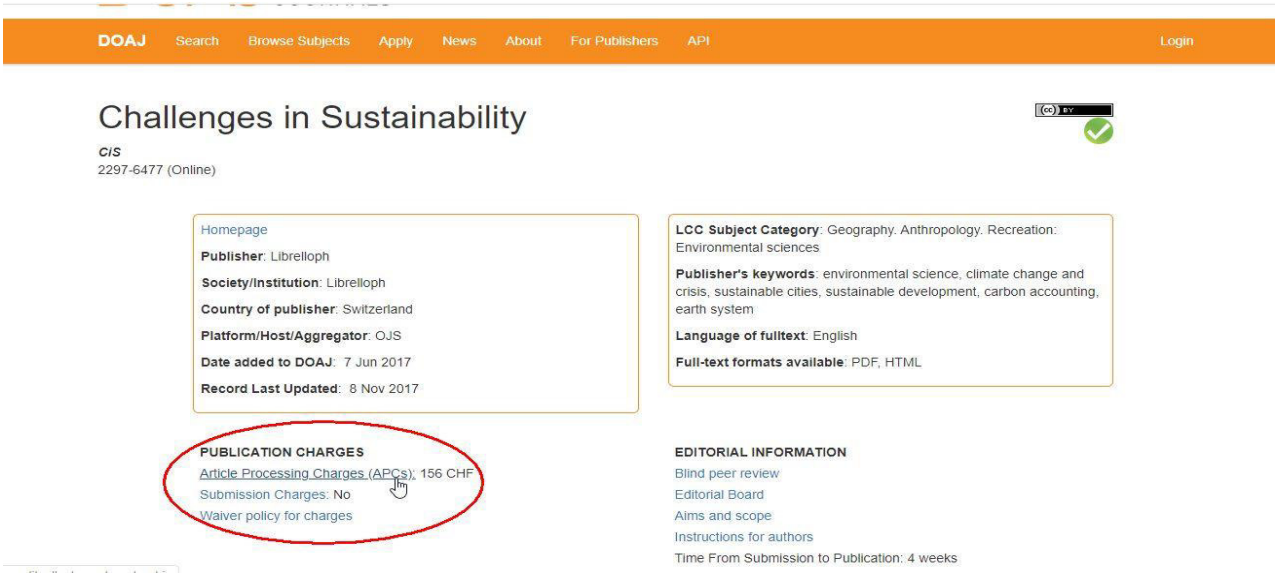
DOAJ was established in 2003 at Lund University, Sweden. Now it contains over 1200 open access journals covering all areas of science, technology, medicine, social science and humanities. DOAJ offers membership in 3 different types: Publisher, Ordinary member, Sponsor.

[2] https://doaj.org/about#aimscope



12. DOAJ homepage

By each record we can find the data about the article, journal, publisher, language, full-text format. The Full Text button guides directly to article, where is available. If we use the filters on left hand side and modify our search only for journal we receive information about it (e.g. when added to DOAJ, publisher, country, ISSN, full-text formats) There is some information about the publication charges. If the publisher requires APCs (Article Process Charge), then it should be indicated.



13. Journal data in DOAJ

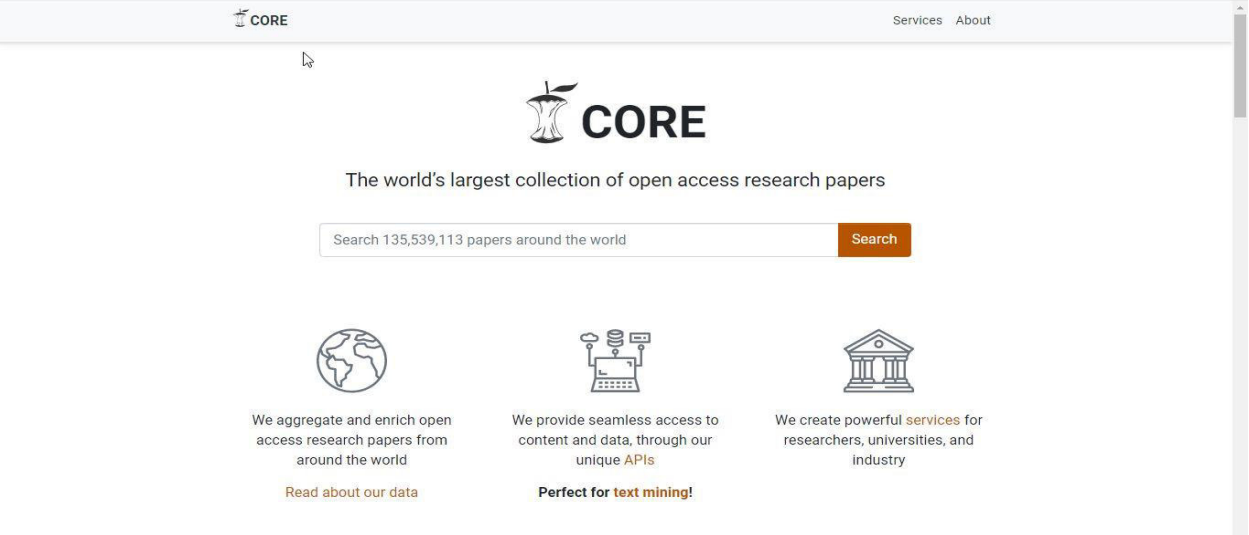
4.3 CORE

CORE is aggregate all open access research outputs from repositories and journals worldwide and make them visibility to the public. CORE supports the right of citizens and general public to access research results; provide easy access to the contents for general public, academic institutions, libraries, software developers and researchers, supports both content producers and consumers; increases the research content using the state-of-the-art technology.

CORE harvests research outputs from data providers worldwide including institutional and subject repositories, open access and hybrid journal publishers.

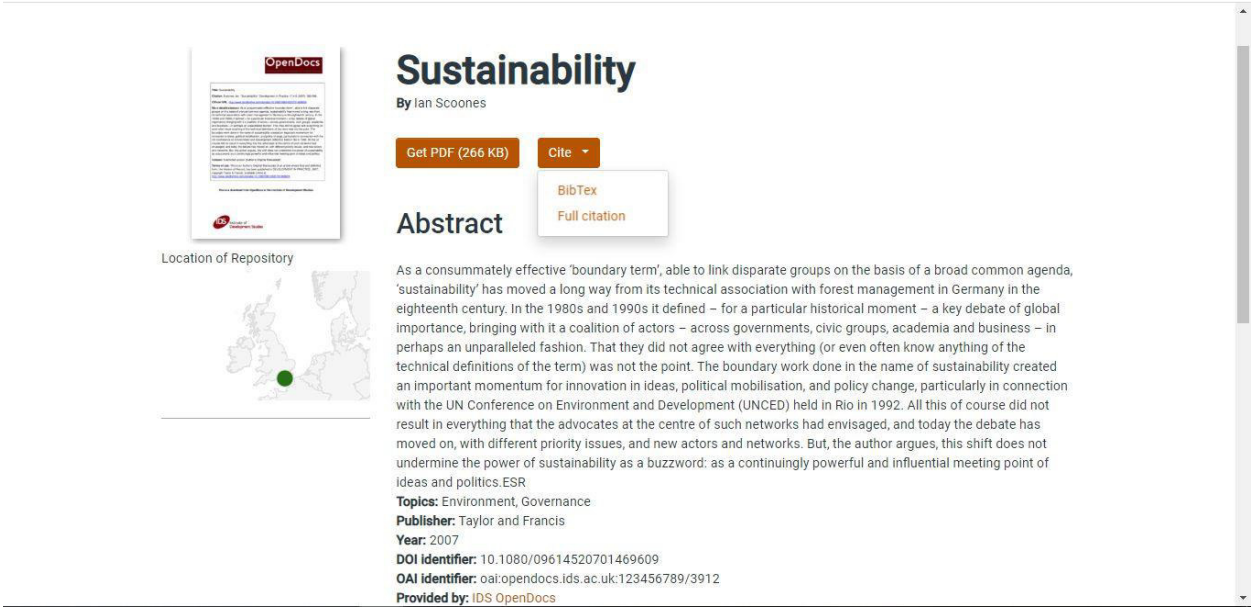
CORE offers several services:

- Access to raw data
    - API for direct access to data,
    - Dataset provides to download CORE data and use in own infrastructure,
    - FastSync helps to keep up to date copy of all CORE data in the infrastructure
- Content discovery
    - Recommender is a plugin for repositories, journals and web interfaces to recommend similar articles
    - Discovery provides some assistance to find freely accessible research papers behind the paywall
- Managing content
    - Repository dashboard is a service for data providers. It offers technical information and statistics to content providers.



14. CORE homepage

CORE can be used easily, after simple search have some options to set up filters (e.g. year, languages, journals, repositories) to refine search. On the search results the records represent the title, author and repository data. Some records can be cited in BibTex or full citation format. Each record represents the location of repository in a map. By the records are suggested similar articles in the same topic.



15. Record in CORE

4.4 OAIster database

OAIster is a catalog of records that represent open access resources. It contains more than 50 million records. This catalog uses OAI-PMH protocol to harvests open access collections.

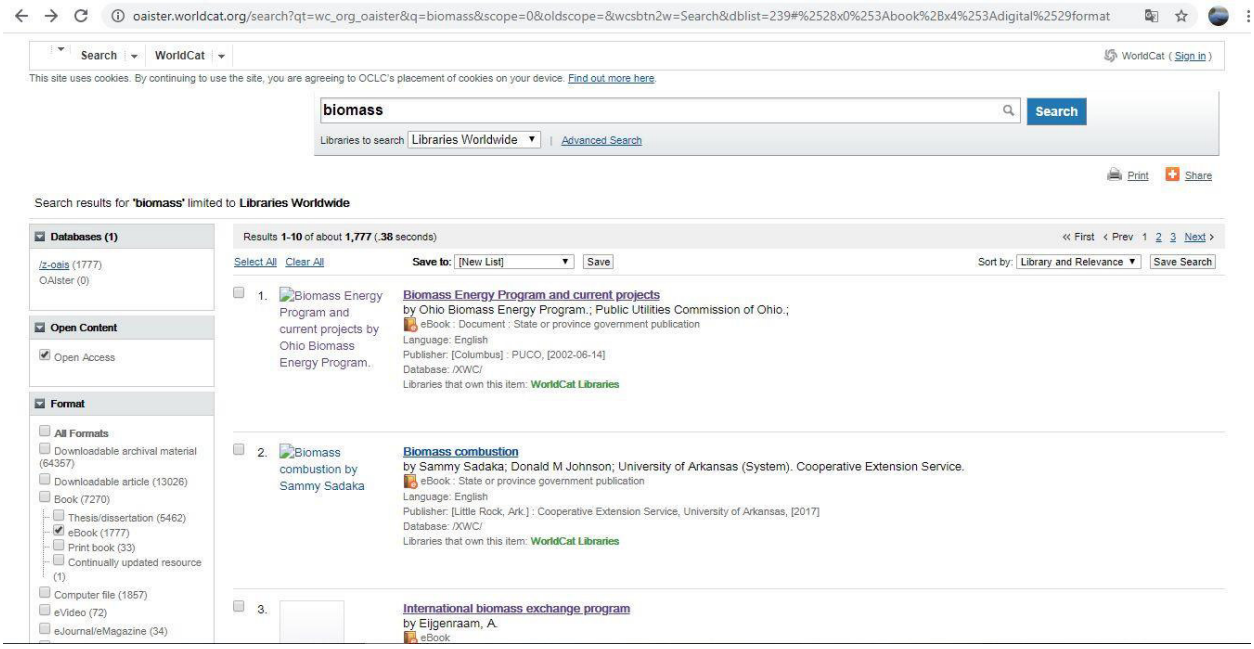
OAIster is a freely accessible site, although the records are available through WorldCat.org. WorldCat Digital Collection Gateway is a service that supports the institutions to add their metadata to the WorldCat. WorldCat Gateway is compatible OAI-compliant repositories therefore they can upload their metadata to OAIster and WorldCat.

OAIster was established by University of Michigan in 2002 to provide digital library resources for public. OCLC formed a partnership with the University of Michigan in 2009 to continue the access to open access collections.

Wide range of materials is available via OAIster:

- digitized books, journal articles, newspapers, manuscripts
- photographic images
- digital texts
- audio files
- video files
- data sets
- theses and research papers.

By the search in OAIster we can sort the search results or refine them. We can select year, languages or formats. Formats assist to reach our purposes of our search. Our example search query is ‘biomass’ and retrieval more than 80000 records that contain the biomass word. We can make favorite list after create a free account. The result of records can be printed and shared via several platforms, applications.



16. Search result in OAIster database

5. Digital Library, Digital Collection

A digital library, digital collection is an online database of digital objects such as image, text, video formats. The materials can be originally in digital formed or digitized items. The digital libraries offer the organizing, storing, searching and retrieving of content. The digital contents can be stored locally or accessed remotely via computer networks. Early history of digital library goes back to 1895 to Paul Outlet and Henri La Fontaine. They thought to collect and make a structure for the world's knowledge catalogue. By the 1980s the electric catalogue known as Open Public Access Catalog (OPAC) was created. An example of early bird of digital library is the ERIC (Education Resources Information Center) that a database of education citations and abstracts from 1969. The term of 'digital library' was generally known in 1994 by NASA/DARPA/NSF.

Advantages of digital libraries:

- **no physical boundary**: the user of digital library should not go to the library in physically. Via internet access the people from over the world can access the contents.
- **round the clock availability**: there is no timetables, it is open 24 hours every day.
- **multiple access**: same contents can be used parallel by numerous institutions and patrons.
- **information retrieval**: the user can use any search term and digital libraries have very user-friendly interfaces
- **preservation and conservation**: digital libraries and collections offer many preservation and conservation concerns that analog formats cannot do.
- **space**: the traditional libraries need physically space for storage, the digital libraries need not.
- **easily accessible and added value**

5.1 World Digital Library (WDL)

The World Digital Library (WDL) was established by U.S. Library of Congress with support of the United Nations Educational, Cultural and Scientific Organization (UNESCO) and in cooperation with libraries, archives, museums, educational institutions, and international organizations from the world. 26 institutions in 19 countries contributed content to the initiative site in 2009. Initiative partners included the national libraries of China, Egypt, Russia, France, United States and several other countries.

The World Digital Library focus on the following principles:

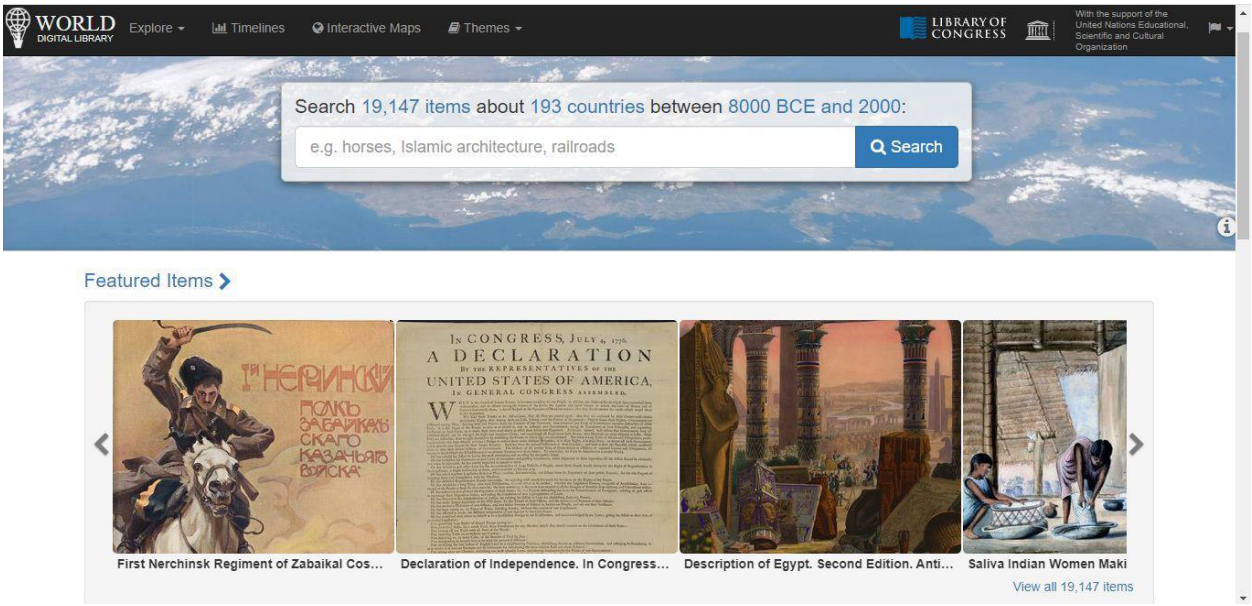
- support the international and intercultural understanding
- spread the variety of cultural content on the Internet
- provide resources for educators, scholars, and general audiences
- build connection among the partners in different countries

Key features:

- **consequent metadata**: each record is described by bibliographical information (metadata). Metadata assist to identify connections between records.
- **multilingualism**: the metadata, supporting content and navigation are translated to 7 languages including Arabic, Chinese, English, French, Portuguese, Russian and Spanish.
- **description**: each record has a description that contents answer the following questions: "What is this item and why is it significant?" This information are written by curators or other expert.
- **digital library technical development**: it uses the state-of-art tools and technologies in cataloguing and website development
- **collaborative network**: it is open in all aspects: access to content, technology transfer among the partners, stakeholders and users.

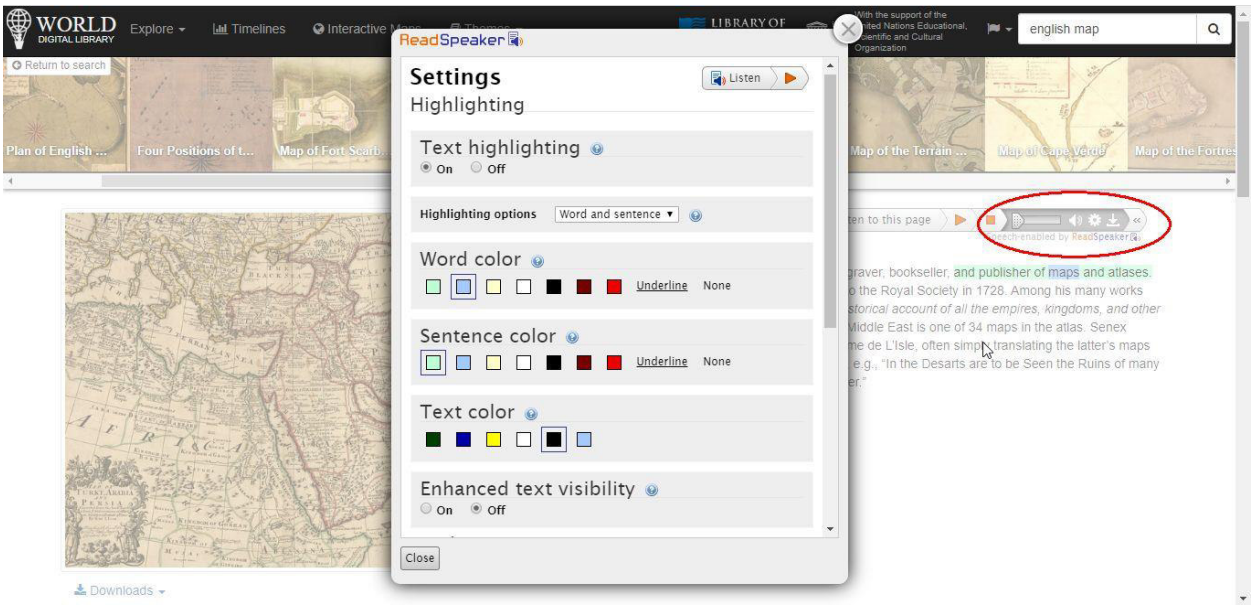
The contents are got into the WDL by digitization centers. The Library of Congress provides training, financial support, equipment, software in Brazil, Russia, Egypt, Iraq and Uganda to establish a digitization centers. The WDL supports and receives content from three digitization

centers: the Iraqi National Library and Archives in Baghdad, the National Library and Archives of Egypt in Cairo, the National Library of Uganda in Kampala.



17. Homepage of World Digital Library

WDL includes books, manuscripts, maps, newspapers, journals, prints and photographs, sound recordings and films. The records can be browsed by place, time, topic, type of item, language and contributing institution. The search term in simple search field can use in metadata, description and full-text in printed books. Each record includes description that explains its significance and historical background. Additional information is provided by curator videos. Other features are available such as image-viewing, timelines, interactive maps. Books, manuscripts and maps on the site are not translated but presented in their original languages. WDL provides a feature 'listen to this page' easily access for blind or partially sighted people and can be downloaded. The 'ReadSpeaker' feature can be set up on demands, therefore can be switch on or off the text highlighting, can be define the word, sentence and text color, and an option is available for speed of speaker (slow, medium or fast).



18. The settings of 'ReadSpeaker' feature in WDL

5.2 EUROPEANA

Europeana is a platform for Cultural Heritage Institutions to showcase their collections alongside other cultural heritage data. It provides access to over 50 million digitized items (books, music, images, maps, artworks). Dedicated thematic collections on art, fashion, music, photography and World War I contain galleries, blogs and exhibitions. Europeana was established by European Commission with support of European Parliament in 2008. At the beginning the most active content provider were France and Koninklijke Bibliotheek in Netherlands. Europeana is financed by the European Union's Connecting Europe Facility and the European Union Member States. The services and website of Europeana is operated by a consortium led by the Europeana Foundation under a service contract with European Commission.

The principles of Europeana were defined by European Commission Recommendation on the digitization and online accessibility of cultural material and digital preservation in 2006.

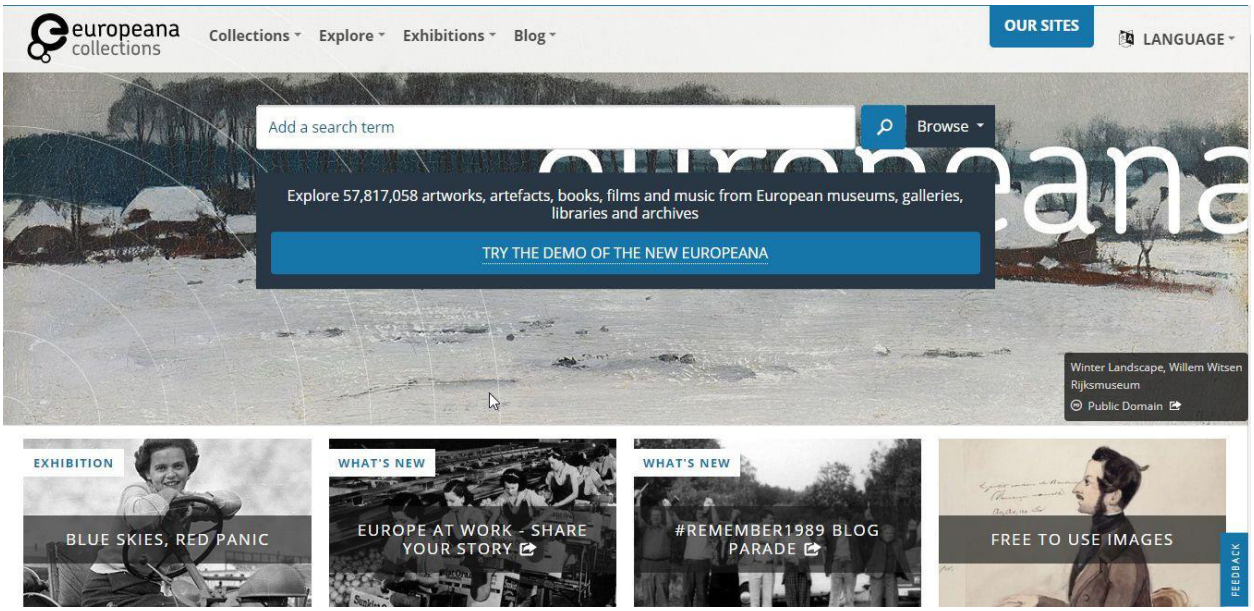
Major principles of Europeana:

- **accessibility**: the material can be easy access for European citizens to use it for leisure, studies or work. The items can be re-used in industries such as tourism and education industry
- **collaborative work**: the digitalization projects should be harmonized together
- **rights**: the intellectual property rights should be harmonized to the certain aspects of copyrights and related rights of European Parliament
- **multilingualism**: a multilingual access point created to search the online digital cultural heritage of Europe
- **preservation purpose**: state-of-art tools and technologies such as web harvesting technique for collecting material from Internet

The Europeana supported several projects and initiation to enhance digitization, technologies and related infrastructures or standard developments. Some significant projects of Europeana:

- **ATHENA (2008-2011)**: this project with 'best practice' initiation has brought the stakeholders and content providers under the roof to evaluate and integrate specific tools
- **European Film Gateway (EFG: 2008-2011)**: unlocked the richness of Europe's film archives for Europeana users and cinema lovers
- **EURO-Photo (2010-2012)**: it has digitized at least 150000 photographs from 10 major European news agencies
- **Europeana Libraries (2011-2012)**: digital collections of 19 of Europe's leading research libraries to Europeana
- **APEx (2012-2015)**: contributes to the development of digital infrastructure which opens up the portal to common cultural heritage of archives
- **LoCloud (2013-2016)**: explores the cloud computing services for small and medium museums
- **Europeana DSI-4 (2018-2020)**: the Digital Service Infrastructure (DSI) showcases and provides online access to Europe's cultural and scientific heritage

- **Europeana Common Culture (2019-2020)**: purposes to develop a harmonized and coordinated environment for Europeana's national aggregators



19. Homepage of Europeana

5.2.1 European Data Model (EDM)

The EDM assists structuring and representing data delivered to Europeana by the variuos contributing cultural heritage institutions. The principles of EDM are based on the core principles and best practice of the semantic web and linked data. Advantages of using EDM:

- single data model bringing together elements of existing data models
- allows to standardize data and use of the fields
- it can be validated against a schema
- allows to use of contextual entities
- profiles and extensions are available for different kind of objects and formats

The Europeana Data Model consists of Core and Contextual classes. The core classes of EDM allow the cultural heritage object to be separated from its digital representation in order for metadata values. These classes represent the core object.

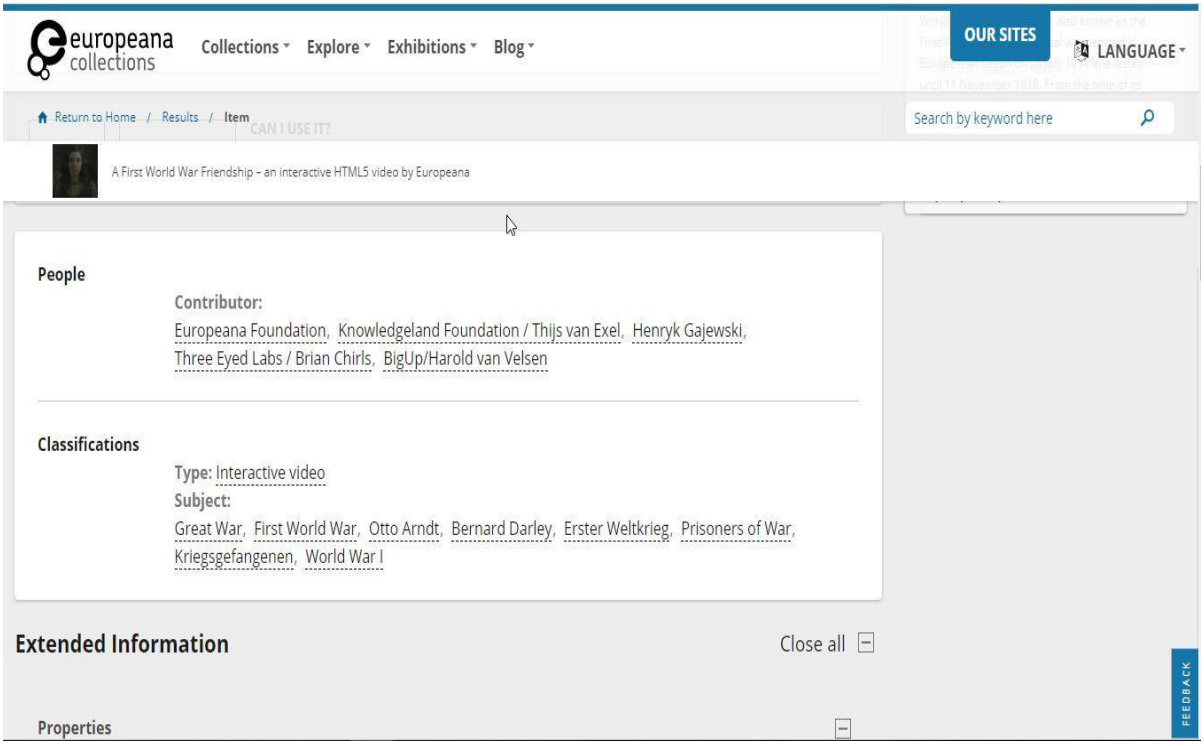
Core classes:

- edm:ProvidedCHO = information on the actual cultural object
- edm:webReaource = the digital representation of the cultural digital object
- ore:Aggregation = the aggregation that groups the two linked together

The contextual classes ensure the 'semantic layer' including concepts from well-established vocabularies like thesauri, authority lists, classifications, content providers, third-party data sources. The contextual classes include 5 classes:

- edm:Agent = who
- edm:Place = where
- edm:TimeSpan = when
- skos:Concept = what
- cc:License = how to use

Sample search of term 'World War I' can be clear the EDM classes and why it is important:



20. Core and contextual classes in EDM with a sample search

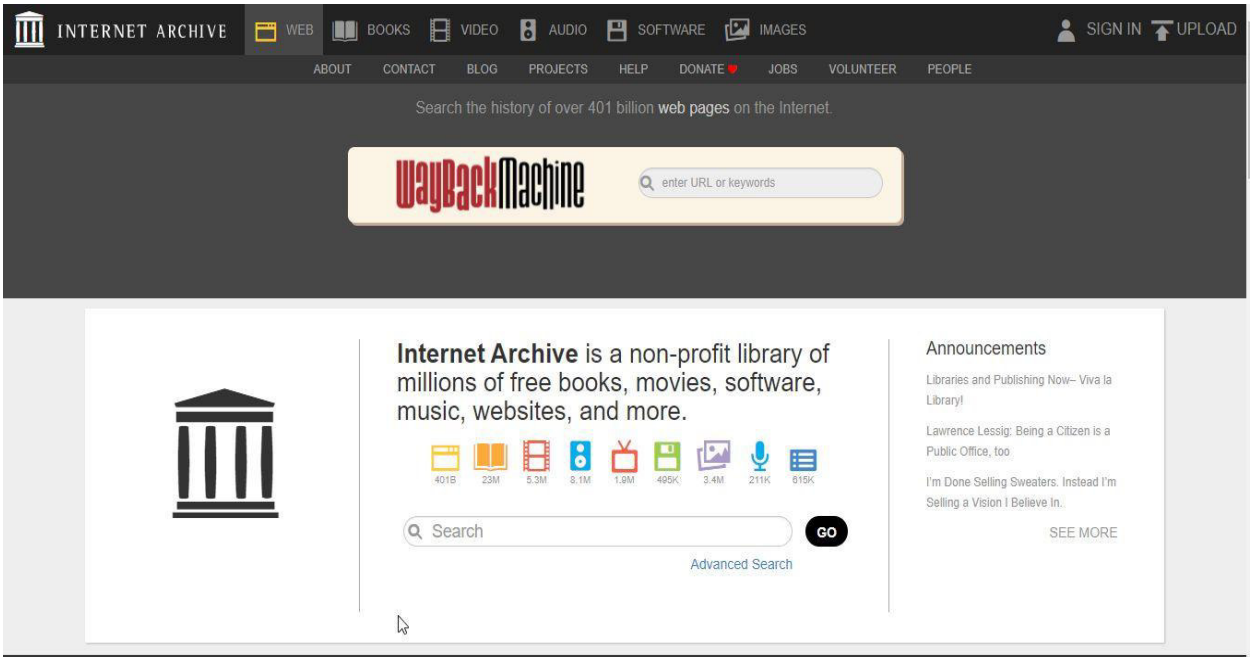
5.3 Internet Archive

Internet Archive is a digital library of Internet sites and other cultural artifacts in digital form. It provides free access to contents for researchers, scholars and general public. The major purpose of it is Universal Access to all Knowledge. The archive process was begun in 1996 by archiving Internet itself through the 'Wayback Machine'. Internet Archive has collaborating work several libraries from over the world and other partners. The 'Archive-It' program provides identify the importance of archive of websites.

The contents of Archive:

- web pages (330 billion)
- books and texts (20 million)
- audio recordings including live concerts (4,5 million)
- videos including TV News programs (4 million)
- images (3 million)
- software programs (200000)

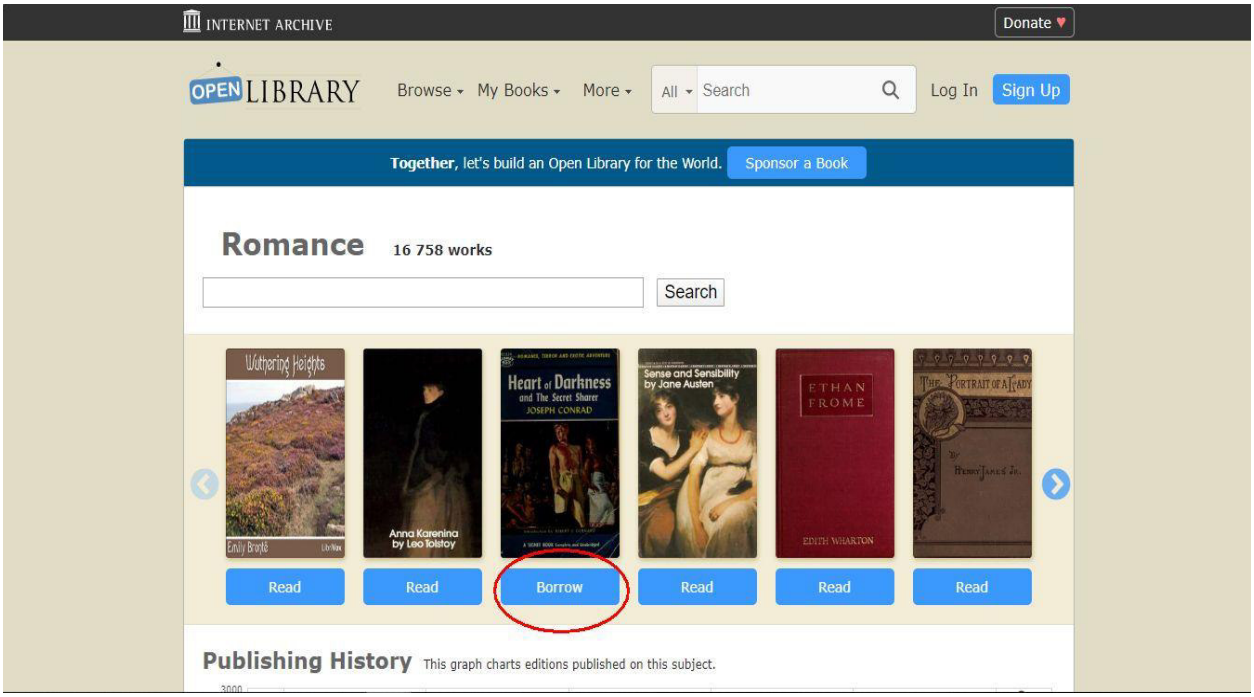
Television programs archiving started in late 2000 with TV news. Internet Archive started to digitize books in 2005.



21. Homepage of Internet Archive

5.3.1 Open Library through the Internet Archive

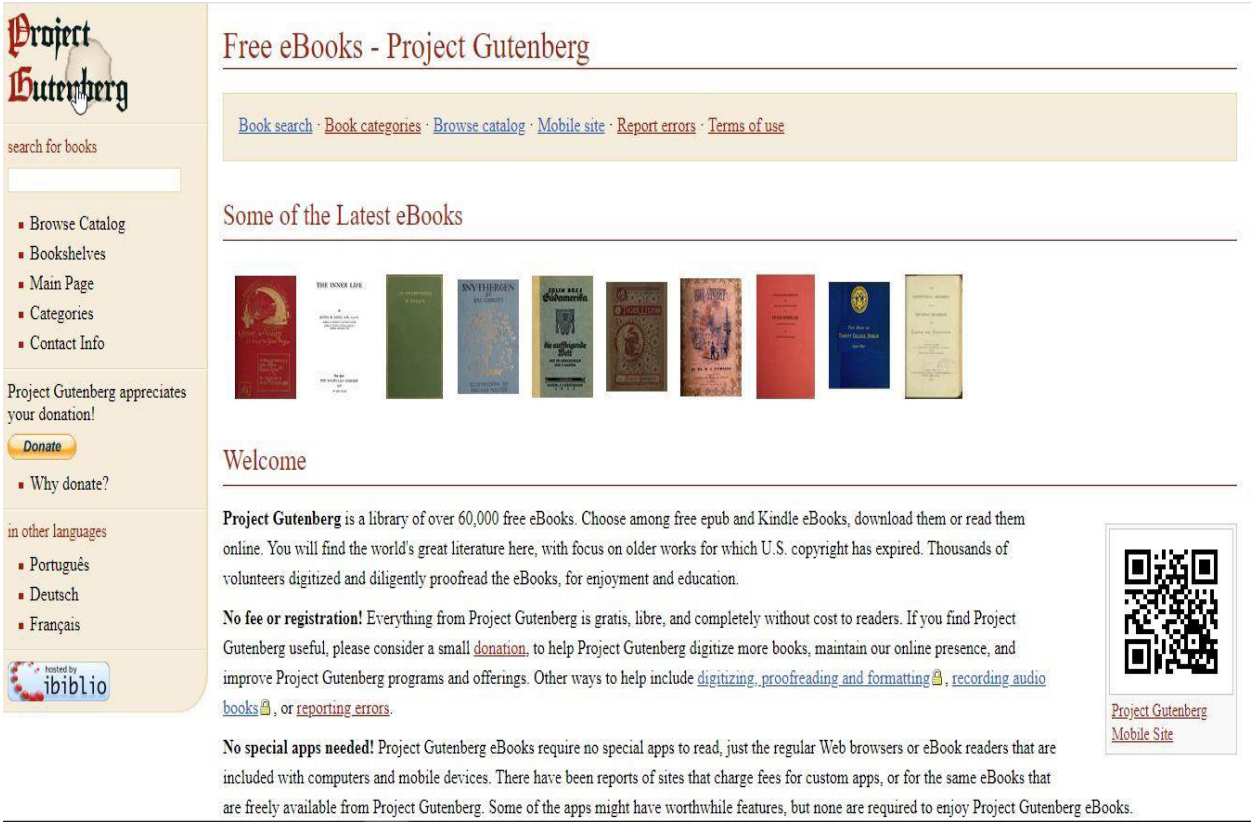
The books printed prior to 1923 can be downloaded. The Archive supports the Open Library initiation, which offers hundreds of thousands of modern books for borrowing. Through the Open Library we can read directly or borrow the books. Books are readable as text or listen as audiobook for assisted the disable people. Open Library offers several topics such as Art, Fantasy, Biographies, Science, Recipes, Romance, Religion, Music, Mystery and Detective Stories, Medicine, Plays, History, Science Fiction, Textbooks and Books for Children. The Open Library initiated a Book Sponsorship Program, which provide a way for readers to contribute preserving of the selected book and become borrowable on Open Library. Before a book is borrowed a personal account should be created and maximum 5 books can be borrowed at once, the loan period is 13 days. Each record is editable by clicking the 'edit' button on the site. If there is no record for a book, the user can create it.



22. Open Library through the Internet Archive

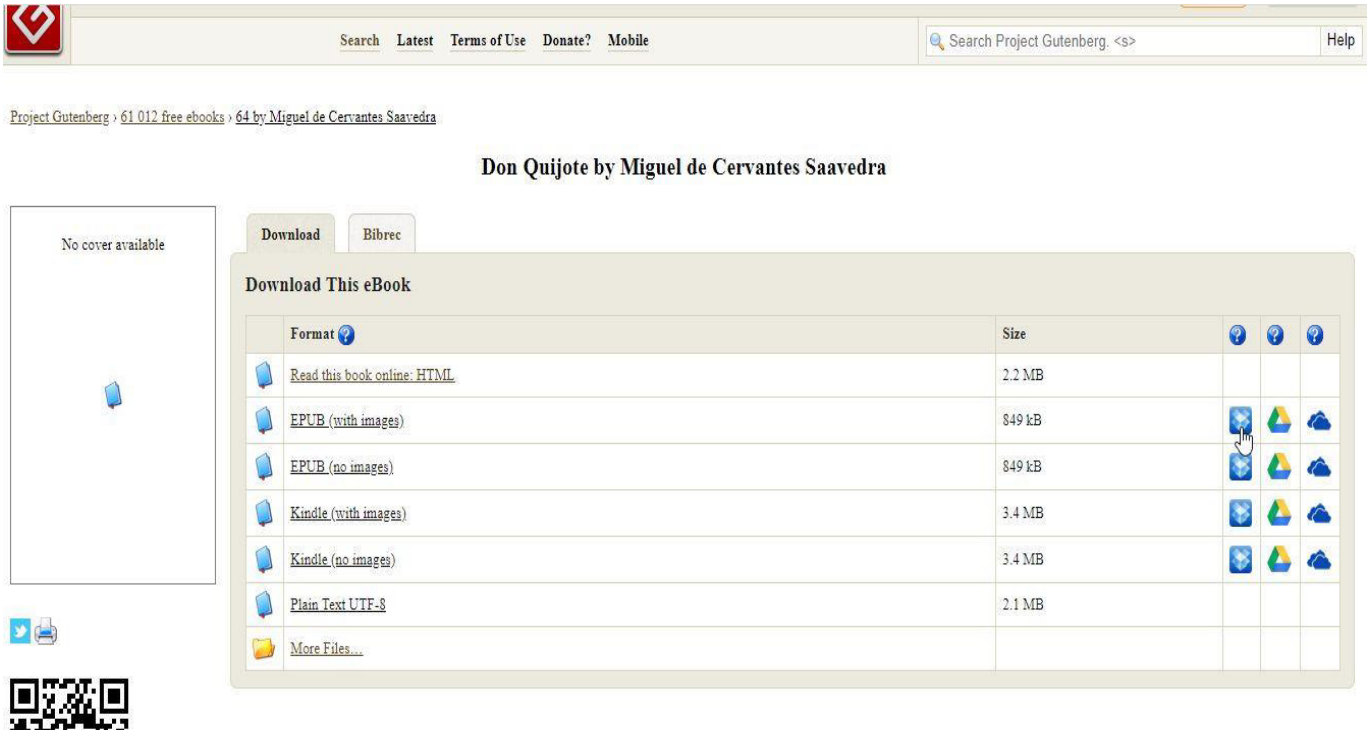
5.4 Project Gutenberg

Project Gutenberg was the first provider of free electronic books. The founder Michael Hart invented eBooks in 1971 and his memory continues to inspire to creation eBooks and related technologies. The major goal of the Project to provide information, books, materials available for general public in forms in vast majority of the computers, programs and people can easily read, use, quote and search. Project Gutenberg is not supported by financial or political power. The co-workers are volunteers, therefore everyone is welcomed to join the Project and contribute it. Everyone is free to do their own E-Books their own way. The E-books in the Project may be freely used in the United States, because are not protected by U.S. copyright law, because their copyrights have expired. They may not be free of copyright in other countries, therefore should check the copyright before downloading or distributing them.



23. Homepage of Project Gutenberg

The site provides search and browse options, to find easily the required item. Several books are available in original language. Downloaded opportunities are numerous, but based on simple text formats for easy access and keep the platform independence. The Project supports the format of Ebook Readers such as Epub and Kindle. By the text downloading it offer to save the selected item to one of our cloud program, such as Dropbox, Drive or OneDrive. The items include bibliographical information also.



24. Downloadable file formats in Project Gutenberg

6. Summary

The libraries and museums from the world play pioneer role in preservation of printed documents and the worldwide accessibility of them. The collaborate work between dedicated librarian staff and the scholars, researchers enhance the efficiency of digitization and online publishing. Popularized open access ways facilitate that digital repositories can open wide door for publicity to achieve more visibility researcher's outputs. Digital collections and libraries contribute to preserving the cultural heritage of countries and enhance the cooperation among them for representing its diversity. Due to digitization and preserving of documents shall be invested capacity into state-of-art technologies and tools and developed standards.

In the last decades numbers of electronic publishing were increased against of the printed documents. More and more periodicals are published online only and demands of customers are increasing for accessing the contents. More people study via e-learning systems therefore shall be developed the interfaces and technical backgrounds. The preserving and distributing of scientific and cultural heritage are based on the collaborate work from over the world.

References, sources

Article Processing Charge. From Wikipedia
https://en.wikipedia.org/wiki/Article_processing_charge (02.10.2019)

Berlin Declaration on Open Access to Knowledge int he Sciences and Humanities
https://openaccess.mpg.de/Berlin-Declaration (02.10.2019)

Brian Lavoie: Meeting the challenges of digital preservation: The OAIS reference model.In: OCLC Newsletter, No. 243:26-30, 2000
https://www.oclc.org/research/publications/library/2000/lavoie-oais.html (15.10.2019)

Clifford A. Lynch: Institutional Repositories: Essential infrastructure for scholarship in Digital Age In: Portal: Libraries and the Academy Vol. 3. No. 2, 2003. pp. 327-336.
https://muse.jhu.edu/article/42865/summary (03.11.2019)

Comission Recommendation of 24 August 2006 on the digitization and online accessibility of cultural material and digital preservation
https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32006H0585&from=HU (09.12.2019)

CORE database
https://core.ac.uk/ (04.11.2019)

Definition of the Europeana Data Model v5.2.8 06/10/2017
https://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Definition_v5.2.8_102017.pdf (09.12.2019)

Desciplinary repositories : list part of the Open Access Directory
http://oad.simmons.edu/oadwiki/Disciplinary_repositories (06.11.2019)

Desciplinary repository. From Wikipedia
https://en.wikipedia.org/wiki/Disciplinary_repository (06.11.2019)

Digital Library. From Wikipedia
https://en.wikipedia.org/wiki/Digital_library (02.12.2019)

A DigiTool rendszer a digitális tartalom menedzselésére. ExLibris.
http://www.ex-lh.hu/site/index.php/HU/digitool (25.10.2019)

Directory of Open Access Books
https://doabooks.org/doab?uiLanguage=en (28.10.2019)

Directory of Open Access Journals
https://doaj.org/about (28.10.2019)

Don Sawyer: The Open Archival Information System and the NSSDC
https://nssdc.gsfc.nasa.gov/nssdc_news/dec00/oais.html (15.10.2019)

DSpace@MIT database
https://dspace.mit.edu/ (28.10.2019)

Electronic Publishing. In: Encyclopedia Britannica
https://www.britannica.com/technology/Internet/Electronic-publishing (06.10.2019)

Európai Digitális Könyvtár. From Wikipedia
https://hu.wikipedia.org/wiki/Eur%C3%B3pai_Digit%C3%A1lis_K%C3%B6nyvt%C3%A1r (08.12.2019)

Europeana.
https://www.europeana.eu/portal/en/ (08.12.2019)

Jihyun Kim: Finding documents in a digital institutional repository: DSpace and Eprints. In: Proceedings in American of the American Society for Information Science and Technology. – Vol. 42 Issue: 1, 2005.19 p.
https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/meet.1450420173 (04.11.2019)

Institutional repository. From Wikipedia
https://en.wikipedia.org/wiki/Institutional_repository (03.11.2019)

Internet Archive

https://archive.org/ (12.12.2019)

Iris Xie, Krystyna K. Matusiak: Digital preservation. In: Discover Digital Libraries : Theory and Practice. – Elsevier, 2016. ISBN 978-0-12-417112-1, pp. 255-279.

ISO 14721:2012 – Space data and information transfer systems – Open archival information system (OAIS) – Reference model. Standard, 2012, 126 p.
https://www.iso.org/standard/57284.html (16.10.2019)

LUTPub

https://lutpub.lut.fi/ (06.11.2019)

Martin Doerr, Stefan Gradmann, Steffen Hennicke, Antoine Isaac, Carlo Meghini, Herbert van de Sompel: The European Data Model (EDM). 149. Information Technology, Catalouging, Classification and Indexing with Knowledge Management. World Library and Information Congress, 76th IFLA General Conference and Assembly. 10-15 August, 2010, Gothenburg, Sweden.
https://www.researchgate.net/publication/303058300_The_Europeana_Data_Model_EDM (10.12.2019)

National Data Repository. From Wikipedia
https://en.wikipedia.org/wiki/National_Data_Repository (10.11.2019)

Óbudai Egyetem Digitális Archívum
http://lib.uni-obuda.hu/node/79 (25.10.2019)

Open Access. From Wikipedia
https://en.wikipedia.org/wiki/Open_access (02.10.2019)

Open Archival Information System. from Wikipedia
https://en.wikipedia.org/wiki/Open_Archival_Information_System (17.10.2019)

Open Archives Initiative Protocol for Metadata Harvaesting. From Wikipedia
https://en.wikipedia.org/wiki/Open_Archives_Initiative_Protocol_for_Metadata_Harvesting (17.10.2019)

Open Library

https://openlibrary.org/ (12.12.2019)

Oxford University Research Archive (ORA)
https://ora.ox.ac.uk/ (04.11.2019)

Project Gutenberg

https://www.gutenberg.org/wiki/Main_Page (13.12.2019)

Real Repository

http://oaikereso.sztaki.hu/kereso/index.php?type=0 (05.11.2019)

Subject Repositorires.
https://utrgv.libguides.com/oa/repositories (06.11.2019)

Tanya Gray Jones, Lucie Burgess, Neil Jefferies, Anusha Ranganathan, Sally Rumsey: Contextual and provenance metadata in the Oxford University Research Archive (ORA). In: Emmanouel Garoufallou, Richard J. Hartley, Panorea Gaitanou (eds.): Metadata and semantic research : 9th research conference, MTSR, 2015 Manchester, UK, September 9-11, 2015 : proceedings. ISBN 978-3-319-24128-9 pp.274-285.
https://books.google.hu/books?id=foGBCgAAQBAJ&pg=PA276&lpg=PA276&dq=ora+repository&source=bl&ots=FaFPJTfCB_&sig=ACfU3U1xhSNdOdRjXAFi7jNHx8fl62CF4A&hl=hu&sa=X&ved=2ahUKEwj4isztgebmAhWIs4sKHeMaDEsQ6AEwCXoECAcQAQ#v=onepage&q=ora%20repository&f=false (04.11.2019)

Vass Johanna (összeáll): Az Europeana születése
https://mandadb.hu/cikk/650499/Az_Europeana_szuletese (08.12.2019)

What is Open Access?
https://www.openaccess.nl/en/what-is-open-access (04.10.2019)

World Digital Library
https://www.wdl.org/en/ (04.12.2019)