

Internet Culturale

[[Cultural Internet]].

Catalogues and Digital Collections from Italian Libraries



Italy
Laura
Ciancio



The Istituto centrale per il catalogo unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU or Centralised Catalogue Institute for Italian Libraries), within the Ministero per i beni e le attività culturali (MiBAC, or Italian Ministry of Cultural Heritage and Activities) carries out the coordination functions of the Servizio Bibliotecario Nazionale (SBN, or National Library Service), while respecting the autonomy of the libraries themselves. This means providing guidelines, training and the publishing of cataloguing standards for the various types of materials held in Italian libraries.

With the aim of increasing knowledge of the bibliographic collections, the ICCU promotes and coordinates a number of specialised databases: Manus for manuscripts, Edit16 for 16th-century Italian publications and a Registry of the libraries spread around the country.

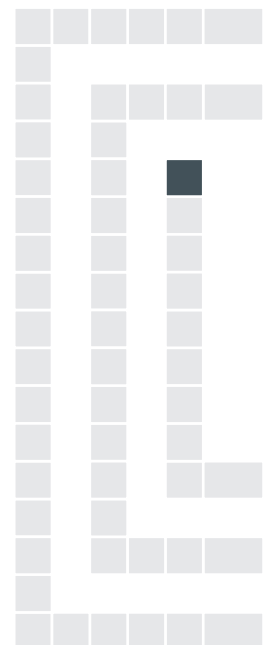
With the arrival of computerisation, ICCU's competencies were extended to this sector, in which it promotes and coordinates activities dealing with the archiving, management, conservation of and access to the digital resources of our bibliographic heritage; it promotes the adoption of standards and implements and coordinates digitisation projects, while undertaking quality monitoring and control.

The Institute manages the *Internet Culturale. Catalogues and Digital Collections from Italian Libraries* portal, which underwent a complete renovation in 2010. The primary aim of *Internet Culturale's* mission is to promote knowledge of the resources of Italian libraries by way of unique access to bibliographic catalogues and by way of access to digital resources, thus enhancing our heritage by offering further cultural study, making use of what had been achieved by the first version of the portal but

with a new goal, that of acquiring both generic users and scholastic users. To this end, an agreement has been reached with the InnoVaScuola portal, an initiative of the Department for Computerisation of the Public Service and Technological Innovation and the Ministry of Education, Universities and Research, allowing that portal to reference a highly educational selection of *Internet Culturale's* contents. Lastly, pursuant to Law no. 4 of 9 January 2004, "Measures to Improve Access by Disabled Users to Computerised Tools", it has been necessary to remove barriers to access.

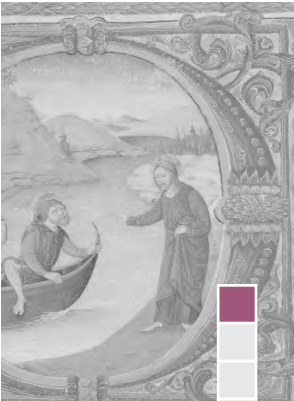
Research on the catalogues: the MetaIndice database.

The centre of the home page displays the form for rapid access to the catalogues. This is a simple search tool which is familiar to users of the web, the default values being those of the



110

The screenshot shows the 'Internet Culturale' website interface. At the top, there's a navigation bar with links like 'Home', 'Esplora', 'Strumenti', 'Partner', 'Chi siamo', and 'Mappa'. Below this is a search bar with the text 'Ricerca: piranesi'. The main content area is titled 'Risultato della ricerca bibliografica' and shows 'Pagina 1 di 80 (799 risultati trovati)'. It lists search results for 'Piranesi, Giambattista' and 'Piranesi nei luoghi di Piranesi'. The interface includes filters for 'Catalogo bibliografico', 'Autore', and 'Varianti al nome'. The bottom of the page shows a footer with the text 'Uncommon Culture'.



Internet Culturale, in its role as aggregator of digital content, aspires to spread the fruits of the activities carried out not only in projects co-financed by the Ministry but by the whole librarian community in Italy.

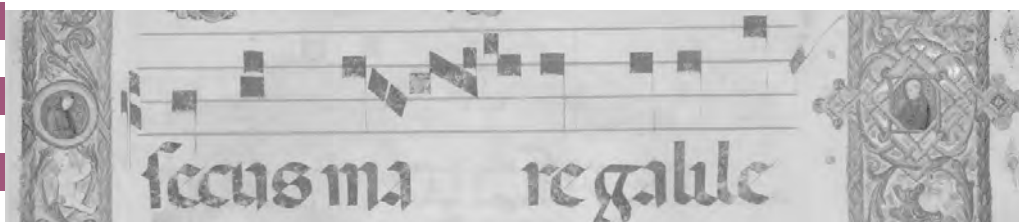
catalogues, and other options allowing separate searches of the Digital Library or the website itself, or requests for advanced searches for the more expert user. Catalogue searches cover the catalogues SBN, Edit16, Manus, the Digital Library, the Rete della musica italiana (ReMI, or Italian Music Network) and the multimedia materials of the portal's CMS. Search results are in summary form but detailed records can be accessed via the original database.

The integration of the bibliographical search into a single interface is done by way of a complex system of indexing of the various databases. This became the most complex feature of the whole project, in part because of the sheer mass of data (extraction of the SBN records alone involved approximately 11 million records) and because of the different characteristics of the original databases and their

export formats. It was necessary to define a common data model, using as the data model's set of descriptive elements the properties and schemas formally defined in the Dublin Core Metadata Terms (DCMT), and to create a system of content uniformity in the form of metadata. Data population is by way of http crawling and file system crawling, and OAI-PMH 2.0. data from the various sources are transformed into the common profile using special plug-ins or XSL-T transformers. The Metaindex thus produced is then updated weekly for the SBN records, while updates to the other databases will take place upon request; at present, data transmission protocols vary according to their specific nature.

The search engine based on Lucene and SOLR open source software creates specialised indices for each of the fields present in the common profile. Searches are made on them, during which a relevancy ranking is established with respect to the request. This ranking is based on standard algorithms using statistical techniques (TF-IDF, cosine metrics, etc.) and may be enhanced by an artificial intelligence mechanism like CBR (Case Based Reasoning) that can keep track of choices previously made by users having similar profiles. For handling advanced enquiries, the search engine supports the traditional Boolean operators (AND, OR, NOT), searching by phrase, and searching by initial graphemes (the start of a written word). In addition, there is support for recursive browsing of content using taxonomies (Dewey browser). Lastly, the engine supports thesauri or ontologies for the semantic expansion of enquiries and the automatic or semi-automatic identification of related terms. This mechanism has been tested for dealing with synonyms and authors' pseudonyms. The presence of facets containing the most significant metadata present in the documents

The screenshot shows the Internet Culturale website interface. At the top, there is a navigation bar with links like Home, Esplora, Strumenti, Partner, Chi siamo, and Mappa. Below this is a search bar with the text 'Ricerca: piranesi'. The main content area displays 'Risultato della ricerca bibliografica' with a list of search results. The first result is for 'Piranesi, Giambattista' and includes a small image of a book cover. The left sidebar contains various filters such as 'Tipo documento', 'Livello bibliografico', 'Autore', and 'Altri nomi'. The bottom of the page shows 'Data editoriali' and a footer with the 'Uncommon Culture' logo.



found by a search allows post facto filtering and refining of the results of a search by combining criteria. In this way, users are given the possibility of not having to define beforehand the criterion to be used but to request only what is of interest (the main terms); in a second step they can then better focus on the results obtained.

The centralised nature of the search services gives uniformity to the user's experience and the faceted presentation makes the results dynamic so that they can be further processed or refreshed. Catalogue search results are in summary form but detailed records can be accessed via the original database.

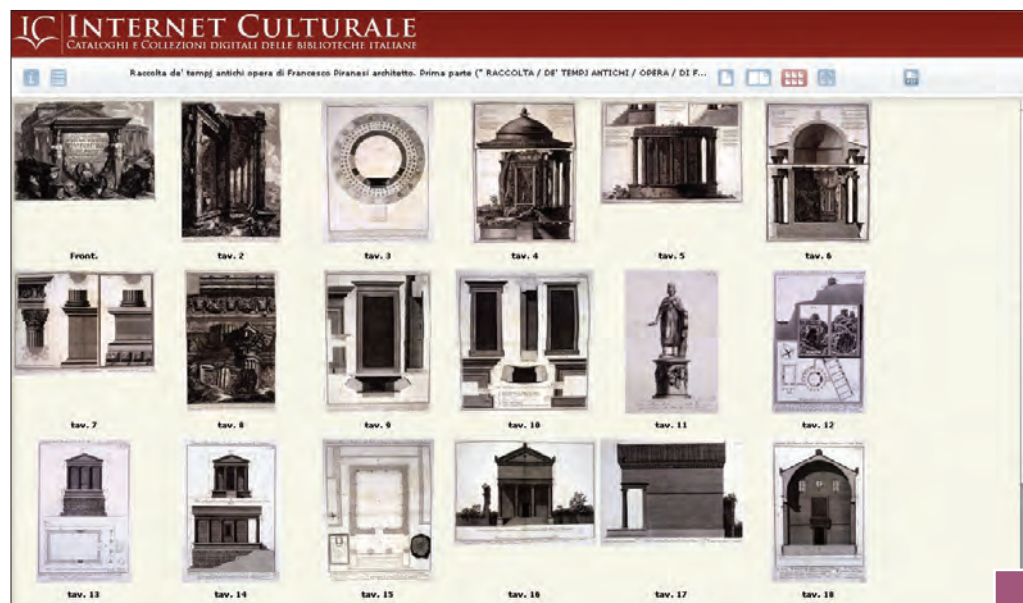
Digital Library: the IndiceMag Database.

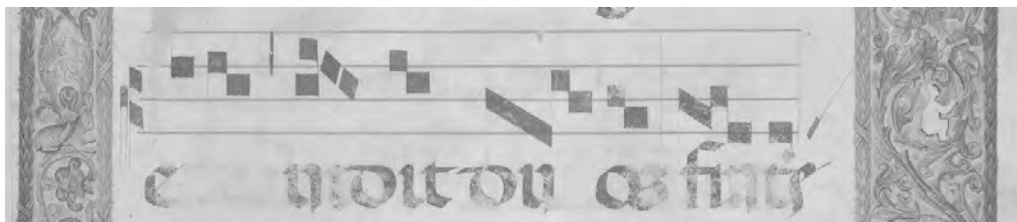
The search form allows enquiries to be made on the computerised resources of the digital collections making up the Digital Library

database. The standard for the metadata that allows the handling of digital objects and that underlies the Digital Library database is the Mag Schema. The metadata, harvested by the portal services from the various digital repositories that are partners of *Internet Culturale*, make up a file that is analogous to MetaIndice and independent of it, to which it passes its own data. The Digital Library's IndiceMag is managed and updated using the OAI-PMH 2.0 protocol, making calls to harvesting services.

In the Digital Library the summary data sheet also includes a preview of the digital resource; this allows direct access to the presentation facility without necessarily going through the detailed record.

The search results are specific to the digital resources; the data sheet returned contains a description of the item, corresponding to the BIB section of the MAG mainly with reference to the analogue data and the specific information regarding the digital resource: repository of



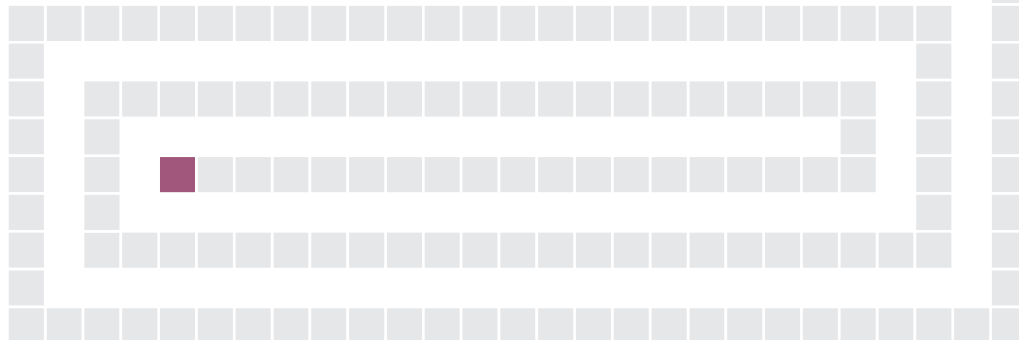


origin, accessibility, degree to which digitisation has been completed, digital format, agency, project.

In addition, in the page showing the details of the digital resource, tools are made available for linking the resource with other resources that could be of interest or have an affinity with this particular resource, thus giving the user the option of exploring the Digital Library using cross-referencing criteria. The main aim of these tools is to make suggestions and stimulate the user's curiosity regarding contents of the Digital Library that may not be immediately obvious. To this end, the detailed data sheets whose documents have a mother-analytical relationship (levels "m" and "a" in the MAG) are highlighted and provided with links. Furthermore, a comparison is made with identifiers of other (Metaindice) databases, the link to the MetaIndice and the item both being shown. There is also a "suggestion" mechanism for consulting other resources based on

the concept of similar documents. The identification of such documents is based both on information of a semantic nature and on statistical algorithms that process a number of metadata fields in order to identify similarities among the keywords. Identification of similarities takes place using the contents of the metadata items dc:title, dc:creator, dc:relation, dc:contributor. Lastly, and still with the aim of providing suggestions as to related resources, the search engine includes techniques of *clustering* (unsupervised statistical grouping of documents) based on the subject of the resource (if present). Such techniques arrange by similarity documents having the same subject by extracting descriptors (themes) of a semantic nature, identifying correlations on a statistical basis, selecting the most significant of them and creating clusters of similar documents. Those descriptors then become suggestions for possible search arguments (themes) and further browsing in the Digital Library.





in order to select the page desired from the previews. Zooming, measuring, rotation and full-page viewing are available; if the owners of the copyright allow it, an extract or the whole resource may be downloaded in .pdf format. The software includes a multi-media player for the portal's media streaming service.

Contents of the portal: the "Esplora" [Explore] page.

The two boxes in the centre of the home page are "Digital Collections" and "Themes", giving direct access to a section of "Esplora" in the menu bar.

"Digital Collections" gathers together summary descriptions of the digital collections indexed in the Digital Library and in MetaIndice. Each summary description carries photographs and links to the fact sheet on the Institute that has created the collection, known as a Partner of *Internet Culturale*, and to the Digital Library which provides answers about the whole collection.

"Themes" is one of the portal's innovations; it is an initiative intended to put on display the numerous multimedia items in "Esplora" which, in addition to the Digital Collections, presents Itineraries, Exhibitions, 3D Pathways, Authors and works. On these multimedia mini-sites *Internet Culturale's* publishing staff has identified portions of autonomous content and for each portion they have produced metadata accompanied by abstracts, subjects and themes, the latter being based on the principles of the Dewey classification system. For instance, the "Itinerario scientifico in Toscana" [Scientific Itinerary in Tuscany] produced 180 records of metadata and themes.

In the left-hand column of each page we find the heading *Eventi e Novità* [Events/What's

New] containing information about activities in Italian libraries. By choice, certain items that were present in the previous version of the portal are no longer present since they were of a more institutional nature, or were like SBN's OPAC services that the user can access in MetaIndice at the time he/she selects an item. From the page "Strumenti" [Tools] it is possible to access the catalogue databases, the names and locations of the libraries, the repositories, the magazine *DigItalia*, and a page of carefully selected links to the resources in the network.

Partners of the portal and *Internet Culturale's* services

Partners, who have their own menu item, are presented using a description of the institution accompanied by photographs. These are institutions that collaborate with the portal in various ways: Regions and Municipalities that co-finance and promote it, research bodies that provide analysis and scientific collaboration, digital consortia, cultural institutions, libraries having their own digital collections. In the descriptions of the partners/institutions, the link *Accesso Patrimonio* [Access to Resources] links to the results of the Digital Library, an overall view of the digital resources of that institution that are present in the Digital Library. In the fact sheets of the collections, too, *Accedi alla collezione* [Access the Collection] gives the results of a search of the collection only in the Digital Library. This arrangement allows libraries and partner institutions of the portal to link to this result through their own institutional sites.

This function was implemented to encourage the participation in the portal of the greatest number possible of libraries, which can carry out digitisation projects using their project

SHORT

ARTICLES

FROM

CLOSED

DOORS

TO OPEN

GATES

funds only for the production of metadata and digital objects, with significant savings in running costs by passing on the results to the service provided by ICCU through the digital repository MagTeca. This service, which has been available since 2005, consists of the free management of low- and medium-resolution digital resources accompanied by MAG metadata, guaranteeing their conservation and dissemination over time through the integrated services of *Internet Culturale*. MagTeca is run on recently re-engineered software based on a "Fedora Commons 3.3" framework.

Joining Internet Culturale

On the strength of its cultural identity as an access portal to the resources of Italian libraries and as a reference point for those with an interest in the world of books, *Internet Culturale*, in its role as aggregator of digital content, aspires to spread the fruits of the activities carried out not only in projects co-financed by the Ministry but by the whole librarian community in Italy.

There are two ways of joining up. Institutions that have a digital repository can distribute their digital resources through the OAI-PMH protocol, making them available to Internet Culturale's harvesting services. Institutions that do not have a digital repository and are undertaking digitisation can consign their digital resources and metadata to the ICCU's MagTeca which communicates with the portal by way of the OAI-PMH protocol.

Technical activities of analysis and mapping need to be coordinated between the portal's services and partner institutions in the case of catalogue databases, to ensure the quality of the metadata both for the acquisition of ICCU in MagTeca and for the harvesting from other

digital repositories on behalf of the portal's services. In the page of the portal dedicated to new partners, guidelines are published on the technical requirements to be respected during digitisation in order to produce the best results in searches and in presentation of content.

Lastly, we wish to point out that those joining the portal are participating in a broader community with resonance beyond their own sector, ensuring their content is made available in the context of *CulturaItalia*, and at an international level on the 'Europeana' portal.

Internet Culturale in numbers

The catalogues:

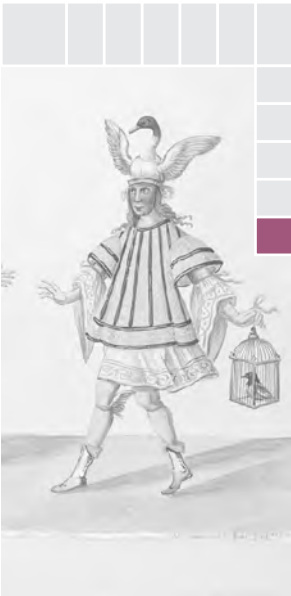
OPAC SBN: **11 million**

Collective catalogue of the libraries of the Ser-



116

Uncommon
Culture



vizio Bibliotecario Nazionale [National Library Service], providing services to the end user; participation by over 3500 public libraries belonging to local councils and universities, as well as public and private institutions operating in various disciplines.

National Census of Sixteenth-Century Italian Publications (Edit16): **60696**

Database of works printed between 1501 and 1600 in Italy in any language or printed abroad in Italian. Also contains authority control information on uniform titles, authors, publishers, printers' marks. There are also images of printers' marks, frontispieces and colophons.

Census of Manuscripts in Italian Libraries (Manus): **263296**

A database of descriptions and partial digitisations of the manuscripts held in Italian

public, ecclesiastical and private libraries, the aim being to identify and catalogue manuscripts in the Latin alphabet produced from the Middle Ages until the present day.

Rete della Musica Italiana (ReMI, or Italian Music Network): **8496**

A database of materials dealing with the music and musical events in the history of Italian opera, a catalogue and digitisation of scores, libretti, letters, sketches, photographs of sets, etc.) as well as integrated computerised media (manuscripts, printed editions, audio).

Digital Library: approx **650000** metadata items; approx 8 million digital files. The Digital Library is the database of administrative and operational metadata (Mag) relating to the digital objects of the partner institutions of *Internet Culturale*. It allows integrated searches to



be made across the data on the various digital resources present in the collections.

Using OAI-PMH, the Digital Library encompasses the following digital repositories:

- BAICR Sistema Cultura, set up in 1991, is a consortium that brings together a number of Italian cultural institutions: of these the Istituto Luigi Sturzo, the Fondazione Lelio e Lisli Basso-Issoco, and the Società Geografica Italiana make their collections available to *Internet Culturale*.

- Biblioteca Italiana (BibIt) curated by the Italian Studies Department of the "La Sapienza" University in Rome, is a digital library of more than 1700 texts representing Italian cultural and literary tradition from the Middle Ages to the Twentieth Century. "Scrittori d'Italia", a series of texts produced, beginning in 1910, by the publisher Laterza of Bari; shortly also "Incunaboli volgari", the digitisation of about 1800 incunabula in the vernacular held in Italian and foreign libraries.

- Biblioteca Laurenziana in Florence, the Plutei collection: 16955 records.

- Biblioteca nazionale Braidense in Milan: Emeroteca Braidense: 265000 records of magazines.

- Biblioteca nazionale centrale in Florence: approx 140000 records of cartography, photography, music, rare manuscripts, Galileo Galilei material, Grand Tour.

- Biblioteca nazionale Marciana in Venice: "Geoweb" with a collection of cartography and graphic materials (Piranesi, Vasi): 27600 records. In a separate collection, musical manuscripts (Scarlatti): 800 records.

- Istituto centrale per i beni sonori e audiovisivi (ICBSA, or Central Institute for Sound and Audiovisual Resources) - Rome: 130000 sound recordings protected by copyright; listening for 30 seconds allowed (public interest access), Italian popular music and classical music.

- Istituzione Casa della musica in Parma: 24000 records of music magazines from the Centro Internazionale di Ricerca sui Periodici Musicali (CIRPeM, or International Centre for Music Magazines Research).

- Museo Galileo: approx. 2000 records of digitised volumes on ancient scientific material.

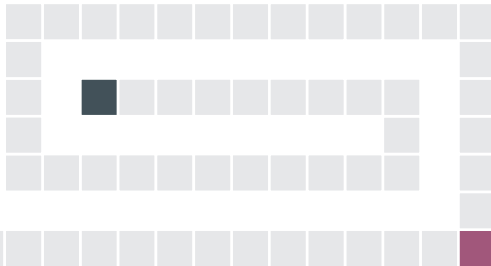
- ICCU's MagTeca: approx. 90000 records, in particular music manuscripts, which manages the digital collections of the following institutions: Conservatorio di musica S. Pietro a Majella in Naples, Conservatorio di musica L. Cherubini in Florence, Fondazione Gioachino Rossini in Pesaro, Biblioteca Angelo Mai in Bergamo (Mayr, Donizetti), Museo Donizettiano in Bergamo, Biblioteca nazionale universitaria in Turin (Vivaldi, Stradella), Biblioteca Estense universitaria in Modena (Stradella), Biblioteca nazionale centrale in Rome, Biblioteca dell'Archiginnasio in Bologna, Museo internazionale della musica in Bologna, Provincia autonoma di Trento—Castello del Buonconsiglio, Biblioteca musicale Abbazia in Montecassino, Biblioteca Oratoriana dei Girolamini in Naples, Biblioteca dell'Accademia filarmonica romana, Biblioteca Augusta in Perugia (corali S. Domenico, Morlacchi), Biblioteca Statale in Lucca (Puccini), Istituto musicale L. Boccherini in Lucca (Puccini). Società internazionale per lo studio del Medioevo latino (SISMEL or International Company for the Study of the Latin Middle Ages) in Florence, Biblioteca Marucelliana in Florence (Mare Magnum), Società internazionale studi francescani (SISF)-Sacro Convento in Assisi, Biblioteca Casanatense in Rome (bandi e bolle pontificie), Biblioteca Alessandrina in Rome (magazines), Biblioteca di storia moderna e contemporanea in Rome, Biblioteca nazionale in Potenza (magazines), Accademia della Crusca in Florence, ICCU (magazines "Pre-unitari"), Museo nazionale del cinema in Turin and Centro sperimentale di cinematografia.





grafia/Cineteca nazionale in Rome (50000 records of the photographer of the stars Angelo Frontoni).

Shortly also: Biblioteca digitale italiana di Matematica (UMI), Biblioteca Palatina in Parma, BEIC in Milan [European Library of Information and Culture].



Conclusions

ICCU is engaged in an ongoing collaboration with libraries and cultural institutions by providing information and recommendations for the implementation of digital projects, monitoring and promoting membership and participation in the portal in order to make Internet Culturale the virtual meeting place of the entire network of Italian libraries. The implementation of the new release is not a point of arrival, but rather a starting point for a service that is more and more targeted to satisfying the needs of users and professionals.

Home Esplora Strumenti Partner Chi siamo Mappa Lingua: It En Fr Es

ICCU INTERNET CULTURALE
CATALOGHI E COLLEZIONI DIGITALI DELLE BIBLIOTECHE ITALIANE

Il mio spazio personale:
nome utente: *****
Registrati ora! | Dimenticai la password?

SBN attualità
10.324.566 di record bibliografici
4.092 biblioteche partner, 62 poli

BDI attualità
4.532.000 dalle collezioni digitali
40 istituzioni partner di progetti
10 digital repository partner

Ricerca: [input field] [search icon] Cataloghi Biblioteca digitale Sito web

EVENTI E NOVITÀ
Nessuna novità
Archivio eventi e novità
Notizie in RSS

COLLEZIONI DIGITALI
[Grid of digital collections]
Tutte le Collezioni

TEMI
Generalità
Filosofia e Psicologia
Scienze sociali
Scienze naturali e Matematica
Tecnologia (Scienze applicate)
Arti
Letteratura
Storia e Geografia

1861 > 2011 >
Una realizzazione:
ICCU
Istituto Centrale
per il Catalogo Unico

cultura
italia
UN PATRIMONIO
DA ESPLORARE

europea
pensare la cultura

innova
scuola