ISS Project: The Integrated Search System in the National Bibliographic Services

Luigi Cerullo(✉)

Istituto Centrale per il Catalogo Unico delle biblioteche italiane e per le informazioni bibliografiche (ICCU), 00185 Rome, RM, Italy
luigi.cerullo@beniculturali.it

Abstract. The project originates from the need to overcome the criticalities of separate search and retrieval platforms for each of the national systems provided by the ICCU and at the same time to "heal the rift" between the National Union Catalogue (SBN) and the Digital Library (Internet Culturale). The actions aimed at rationalizing the retrieval model of ICCU systems, ensuring a consistent recall of information objects through its access interfaces, go towards two development lines: (1) use of one software platform that is the base for the application of the new Integrated Search System (ISS); this solution allows the creation of a single access point configured as a General Catalogue and, at the same time, the decommissioning of different OPACs whose peculiarities will be implemented in the single platform through dedicated search indexes; (2) native integration between bibliographic records and digital copies of publications through the integration of digital library system's services and the SBN library management software.

Keywords: Collective catalogue · Digital library · Access interface
Management metadata · REST services · Searching and retrieval methods

1 Single Access and Retrieval Platform

1.1 General Goals of the Project

The main goal of the SRI project is to build a distributed information architecture that allows the creation of a single access point to the resources described in the main "management" databases, whose care depends directly and indirectly on ICCU (Edit16 [1], Manus On Line [2], The digital resource aggregator-index [3] and SBN [4]). This access interface or access point must be based on a retrieval model that takes into account the ontological level of the bibliographic universe (resources and related entities that are the subject of different descriptions or information items stored in the various management data-bases). The ontological coherence of the representation can be ensured only by acting directly in the bibliographic information creation phase (in the record making phase), ensuring the presence of identifiers and linking-keys between the different information objects - objects that are different for both data quality and data model. In fact, these objects are created in the specialist databases (Edit16 and Manus), are stored in the digital resource index (Internet Culturale) and are shared in the

G. Serra and C. Tasso (Eds.): IRCDL 2018, CCIS 806, pp. 219–224, 2018.
https://doi.org/10.1007/978-3-319-73165-0_22

Collective Catalogue (SBN), but often refer to the same resources and entities. The actions aimed at rationalizing the retrieval model of ICCU systems are listed below:

(a) The use of a single software platform that is the basis of the new Integrated Search System (ISS) that allows the creation of a single access point configured as a general catalogue, and the decommissioning of the current infrastructure of separate OPACs, whose specifics functionalities will be replicated in the single platform through dedicated search indexes.
(b) Native integration between bibliographic records and digital copies of publications through the integration of digital library system's services with the most used SBN library management software that enables the SBN local nodes to supply the Index of digital resources (Internet Culturale) through automated generation of management metadata.

1.2 Integrated Search System and General Catalogue

The ISS is, at the application level, a single platform that includes a unified-index, meant as unified search profile. This solution allows a configuration of (a) search and retrieval interface (search interface, results lists and contextual search filters) of a single access point, presented as a general catalogue, and (b) individual search databases represented by the dedicated indexes Edit16, Manus and Internet Culturale, all these systems accessible through dedicated retrieval interfaces. The index of the General Catalogue is represented by the convergence of the SBN Unimarc profile and the Manus TEI-MS [5] profile, enriched by information of the digital copies described by the MAG [6] metadata referring to the analogic resource that has been described in SBN, in Manus or in Edit16 (1.3 sub) (Fig. 1).

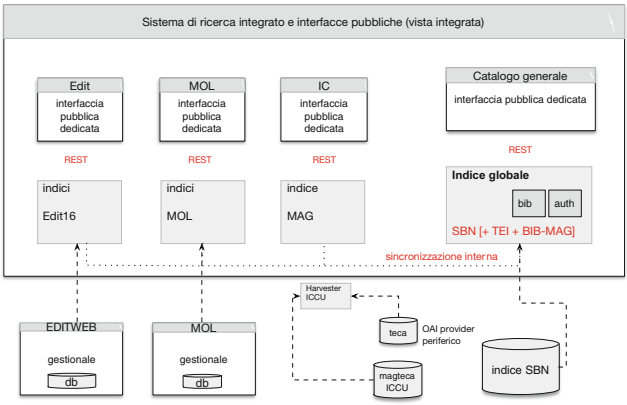

Fig. 1. In the scenery described, public retrieval interfaces represent the specialized sections (dedicated page-types) of a single web platform of fruition (Frontend Web-Applications). Specific search paths - in support of each specialized interface - will be included in the same database search server. The described layout provides that SBN's database functions as a basis for the unified profile.

Diagram of logical relationships between search database records. The diagram in Fig. 2 illustrates the relationships guaranteed in the indexing process of the unified-index.

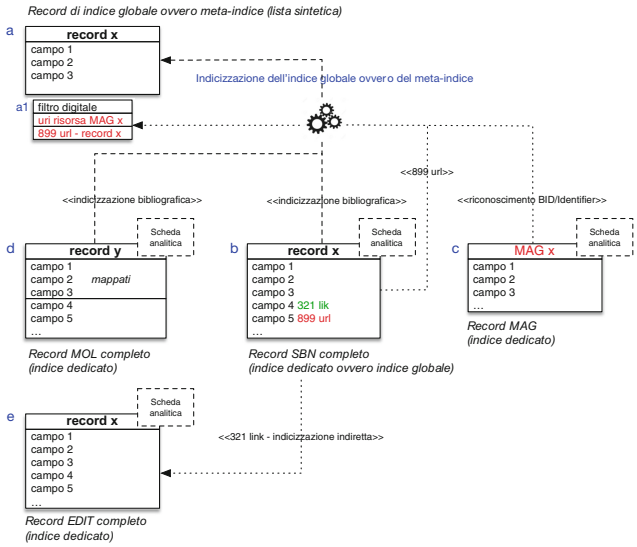

Fig. 2. (a+a1) lists the set of fields of the indices that configure the search masks and the unified-index results lists. For a given record x (e.g. a SBN resource) represented in the unified-index result set, the recognition procedure of a BID that is described into *dc:identifier* tags of indexed MAG (c) will populate index values that serve as filter (a1). This index contextually detect the presence of tags 899$u. The index profile (d) is able to represent TEI-MS information as a whole. Only a subset of these indices will be mapped to the unified-index profile. The index profile (e) represents the EDIT16 specialist search base that is not involved in the unified-index indexing phase. Each record of EDIT16 (here the bibliographic set is exemplified but the logic model is also valid for other entities related to the bibliographic record) is systematically invoked by links in the SBN index records (Unimarc tag 321 as a link to specialized repertories) created in the record-making phase. The index profile (b) represents the set of fields able to represent all SBN's Unimarc information. This information is the core of the integrated representation.

Re-engineering of specialized systems Edit16 and Manus as Client SBN-Marc. The cluster of records (Fig. 2) is made possible by two lines of development: (1) the re-engineering of the management data bases Edit16 and Manus-on-line configured as SBN clients (through the services of its application protocol SBNMarc [7]) and (2) the extension of functional architectures - intended to SBN Nodes - through which the integrated representation of digital resources (represented as bibliographic record attachments) is completed (1.3 sub). These environments will retain their specialist features and will be configured at the same time as Client SBNMarc (also thanks to the reuse of already available technical platforms) that can share all or part of the bibliographic information with the reference system (SBN), forming since the record-making phase the cluster of

distributed records, then coherently represented by the SRI retrieval machine (through its single access point).

Digital as a filter. In the architecture of the reengineered information system, based on the ISS, the digital index of Internet Culturale is not considered as a specialized data base. MAG records are largely de-structured representations of the same resources described in the SBN Catalogue or in Manus and Edit16. The configuration of the general catalogue indices would allow their registration as pure association with the records of the General Catalogue through the development of a unique ID recognition procedure that would come into play during the unified-index indexing phase. This architecture aim to populate specific indexes in order to preserve information about digital content, determined both by the presence of URIs associated in the field of SBN localizations and by the presence of the MAG record referring the same resource (compare diagram of logical relationships in Fig. 2).

1.3 Completion of Integrated Representation: Application Architecture for Peripheral Systems

The application architecture, planned for SBN local nodes, provides full integration of a digital library system's services with most used SBN library management software, through the development of batch generation services of management metadata, based on validated ICCU's mappings and wired in proposed systems. The designed architecture would allow to keep the records in the bibliographic catalogue (in shared environment) aligned with the management metadata (exhibited by the OI-Provider, component integrated in the system) and allow to describe the subset of records with digital attachments resident in the integrated digital library system. The MAG data-set or METS **[**8**]** data-set generated on the basis of batch procedures would be overwritten periodically.

Alignment over time. The functional diagram shown in Fig. 3, explains the solution destined to peripheral systems, together with the perspective of the development of an interface of access and retrieve with integrated unified-index: it will be developed in a long term plan in order to implement the integration-as-alignment between digital objects (represented by MAG or METS data-sets exhibited through OAI-PMH [9] Provider Systems) and bibliographic records of the collective catalogue (SBN), referring to the same resources. This alignment, guaranteed through overwrite cycles and exposure of management metadata, guaranteed on the basis of the implemented logic, would be represented in the central system (i.e. in the ISS, at the unified-index level) by the systematic retrieval of the reference to the digital object present in the single Index of digital resources.

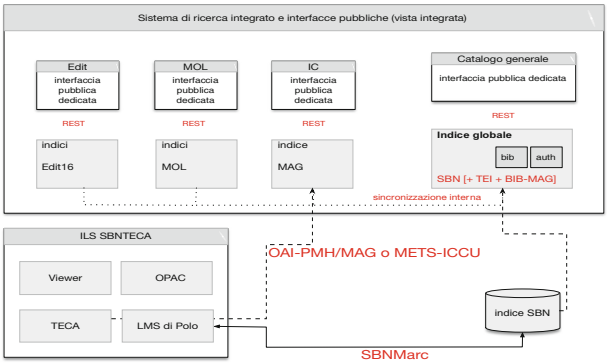

Fig. 3. Architecture of SBN nodes and integration with ISS. In this scenario, through an integrated management environment, the SBN nodes would also become content providers of Internet Cultural, represented - in the Integrated Search System - by one of its indexes: what we define in this work is a single index of digital resources.

The ISS unified-index is represented by the integration of the Unimarc profile of the SBN Index and the TEI-MS profile of the Manus system, enriched by information on the presence of digital copies described by MAG or METS records referring to the same resources described in SBN Index or in Manus and Edit16 data-bases. At first - based on the proposed architecture - only a subset of records that are present in the Single Digital Index (formerly the Internet Culture) would be retrieved, because they relate to digitized resources, mapped - through identifiers - with documents described in the Collective Catalogue. Over time - thanks to the diffusion of the functional architectures proposed in this chapter - the gap between the resources described in SBN catalogue (or within specialized systems) with digital attachments and that can be mapped with digital resources simultaneously described within the Single Digital Index, would gradually decrease.

1.4 Frontend Web Applications (Portals and Search Interfaces)

The Integrated Search System is represented, at the application level, by a single platform that communicates through REST services with Frontend web-applications that represent concretely the public access and return interfaces that the user can use (Fig. 4). These interfaces can be installed either in a single CMS, through which it will be possible to create the Portal of the General Catalogue, both in multiple instances of the same CMS. Each of them represent the service site of each specialized project (Edit16, Manus On Line, SBN, Internet Culturale). The main portal and portals dedicated to specialized databases would share the same template in order to maintain a "family area" according to the models already adopted in the most European projects.

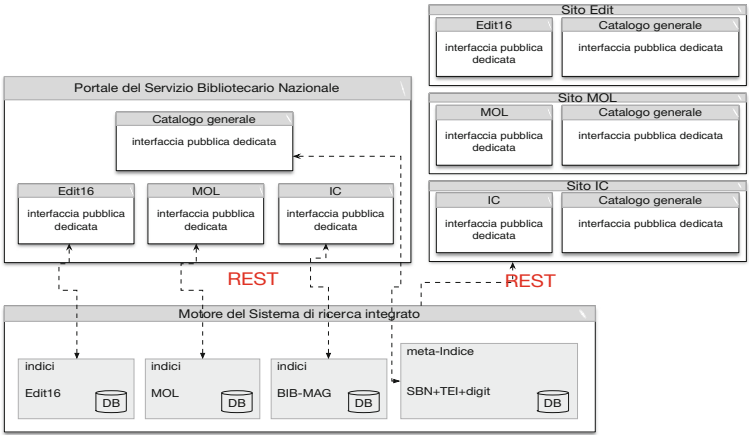

Fig. 4. Dynamic channels, represented by search maps, result-sets equipped with filtering and sorting tools, etc., are considered as "completed" functional components that work as plug-ins installed in CMS, on which the communication portals (or accompanying sites) are based. Avoid using the word "OPAC" to avoid confusion with self-consistent software, at the application level (with its own DB and administration area). The conceptual design of the application platform, based on the separation between Backend and Frontend model - communicating through REST services -, "reduces" the Frontend of the Integrated Search System to a set of components (4 components: Search Interface of General catalogue as a unified-index based on the SBN bibliographic profile, Edit16 Search Interface, Manus Search Interface, and Search Interface of MAG Index), which are deployed on multiple platforms, in different combinations depending on the use case.

References

1. Edit16, catalogue of italian editions of the 16th century. http://edit16.iccu.sbn.it/web_iccu/ihome.htm
2. Manus On Line (MOL), census of manuscripts held by Italian libraries. https://manus.iccu.sbn.it/
3. Internet Culturale, national digital resource aggregator. http://www.internetculturale.it/opencms/opencms/it/
4. Servizio Bibliotecario Nazionale, collective catalogue powered by peripheral catalogs through the services of the SBNMarc application protocol. http://www.iccu.sbn.it/opencms/opencms/it/main/sbn/index.html;jsessionid=E8E3A486F0E3C1F1A7F950414700E177
5. TEI-MS Profile of the Manus On Line System. https://jtei.revues.org/1054
6. MAG Schema. http://www.iccu.sbn.it/opencms/opencms/documenti/mag2-2006.html
7. Technical Documentation and Schema of the SBNMarc Application Protocol. http://www.iccu.sbn.it/opencms/opencms/it/main/sbn/evoluz_indice_sbn/pagina_143.html
8. Documentation on METS metadata framework. http://www.loc.gov/standards/mets/
9. OAI-PMH schema and protocol. https://www.openarchives.org/pmh/