

SIMONETTA BUTTÒ

Gli archivi sonori al crocevia tra scienze fonetiche, informatica umanistica e patrimonio digitale

La sempre crescente attenzione che il mondo della ricerca scientifica riserva al patrimonio immateriale ha contribuito a evidenziare la grande varietà di archivi sonori, realizzati nel corso degli anni, che riguardano molti campi delle scienze umanistiche e sociali: dalle discipline linguistiche (fonetica, fonologia, dialettologia, sociolinguistica) alle scienze che si occupano della voce, dell'elaborazione del linguaggio naturale, del riconoscimento automatico del discorso, fino alla storia orale e all'etnografia, alla sociologia, all'antropologia.

Questi archivi contengono da una parte una grande quantità di dati di ricerca, e rappresentano dunque fonti importanti per le scienze sociali e le discipline umanistiche, ma, allo stesso tempo, costituiscono ormai anche un settore ben definito del patrimonio culturale del Novecento.

Quello della loro descrizione, conservazione e valorizzazione, in Italia e all'estero, è un tema di grande attualità, che acquista dimensioni sempre più estese con il progredire delle soluzioni tecnologiche applicabili ai processi di acquisizione dei dati, alla loro consultazione e all'utilizzazione nei lavori di ricerca. Sono oggi sempre più numerosi gli istituti universitari, le biblioteche e i centri culturali che raccolgono e custodiscono documenti sonori, consentendone la consultazione nelle rispettive sedi o nella rete. Tuttavia, come accade per tanti altri tipi di risorse digitali prodotti dalle istituzioni, numerosi "giacimenti" di documenti sonori sono ancora da scoprire e molti di essi sono poco conosciuti e scarsamente utilizzati: il fenomeno della dispersione li riguarda molto da vicino, anche a causa della loro varietà tipologica e della delicatezza delle problematiche relative alla diffusione in rete di voci e di suoni.

Ai tanti interrogativi posti in questo Convegno in merito al censimento, alla descrizione, alla diffusione, all'uso e al riuso ai fini della ricerca di un tale patrimonio di conoscenze trasversali, è forse possibile fornire alcune risposte prendendo spunto dall'esperienza che l'Istituto Centrale per il Catalogo Unico delle biblioteche italiane-ICCU ha sviluppato all'interno delle reti istituzionali italiane ed europee, che coordina o alle quali ha partecipato fin dalla sua istituzione, con l'obiettivo di diffondere procedure condivise, strategie comuni, flussi di lavoro standardizzati.

Un quadro tanto eterogeneo non deve spaventare, deve semmai indurre a ragionare sulla necessità di adottare standard riconosciuti da un'ampia platea, anche internazionale, a garanzia della piena interoperabilità tramite protocolli di scambio.

Non è più tempo di andare ciascuno per conto proprio, ora meno che mai, perché è diventato indispensabile razionalizzare le risorse di personale, scarse ovunque,

e quelle economiche, sempre incerte, ma anche perché le tecnologie che possediamo ce lo consentono con una facilità prima impensabile.

In questo contesto, dunque, la parola d'ordine non può che essere “condivisione”, lo stesso principio fondante della comunità creata oltre trenta anni fa dal Servizio bibliotecario nazionale-SBN, che oggi accoglie oltre 6500 istituzioni culturali di ogni tipologia funzionale e di ogni appartenenza amministrativa che partecipano alla catalogazione collettiva del patrimonio bibliografico italiano¹.

Oggi SBN è ben di più di un base di dati bibliografica: nei suoi 18 milioni di registrazioni (con oltre 98 milioni di esemplari localizzati in Italia) sono presenti vari materiali non librari, sicuramente di largo interesse per il pubblico generale, ma anche per gli studiosi: ritratti, immagini di luoghi (mappe, stampe, fotografie, cartoline, ecc.), autografi, carteggi o documenti riferibili a persone, mostre virtuali, oggetti d'arte, cimeli e – ormai non più trascurabili per qualità e quantità – risorse audio o video relative sia alle persone (interviste, discorsi registrati) che a luoghi (documentari), oltre ai cosiddetti “materiali non pubblicati”, come conferenze, appunti, dispense, esecuzioni musicali presentate nei festival o in piazza, oppure per esami di concorso.

SBN ha infatti alle spalle una lunga tradizione nell'acquisizione a catalogo di documenti sonori o musicali: l'interesse per la documentazione musicale è nato negli anni Ottanta del Novecento, con la creazione di una apposita *Base dati Musica*, e nel corso degli anni sono state prodotte linee guida per la catalogazione in SBN della musica, dei libretti, delle registrazioni sonore, dei video e delle risorse elettroniche digitali², fino alle recenti *Norme di catalogazione delle risorse musicali non pubblicate* (maggio 2018), tutte consultabili su piattaforma wiki all'interno del sito dell'Istituto³.

Un grande traguardo per l'implementazione dei beni sonori e audiovisivi nel catalogo, arricchiti da link alla corrispondente risorsa digitale, se disponibile, è stato raggiunto alla fine del 2015 grazie al riversamento dei record bibliografici dell'Istituto centrale beni sonori e audiovisivi (ICBSA): operazione alquanto complessa, per compiere la quale sono state messe in campo le migliori energie dei due istituti centrali, realizzando una sinergia tale da rendere sostenibili anche le operazioni di pulizia del catalogo dalle duplicazioni, inevitabili nei processi di importazione massiva di tale portata.

In SBN hanno dunque fatto il loro ingresso le registrazioni su supporto audio e video acquisite dall'ICBSA per deposito legale: circa 300.000 monografie (dischi in vinile, audio e videocassette, compact disc, DVD, nastri, ecc.), e 1.500.000 spogli, ovvero brani sonori (e relativi link alla copia digitale). In questo prezioso insieme sono incluse sia fonti musicali, sia fonti orali, come ad esempio l'Archivio Etnico Linguistico Musicale - AELM⁴.

¹ <https://opac.sbn.it/opacsbn/opac/iccu/free.jsp>.

² https://www.iccu.sbn.it/export/sites/iccu/documenti/2012/Guida_alla_catalogazione_in_SBN_Musica.pdf.

³ http://norme.iccu.sbn.it/index.php?title=Norme_musica_non_publicata.

⁴ <http://www.icbsa.it/index.php?it/116/aelm-archivio-etnico-linguistico-musicale>.

Frutto dell'importazione di tali materiali è la possibilità, dal 2017, di visualizzare oltre 200.000 incipit musicali con il pentagramma e le note, che attraverso l'OPAC di Polo SBN Web possono essere anche ascoltati.

Per arricchire ulteriormente la presenza delle risorse musicali nel catalogo nazionale, nell'estate del 2016 è stato firmato un Accordo tra ICCU, Internationales Quellenlexikon der Musik e.V. (RISM) e Associazione Italiana delle Biblioteche, Archivi e Centri di documentazione musicale (IAML-Italia)⁵, con l'obiettivo di individuare e importare nel catalogo SBN i record relativi a risorse conservate in biblioteche italiane oggi presenti solo nei cataloghi cartacei ed elettronici del RISM; rendere i record bibliografici e di autorità presenti in SBN ricercabili e indicizzabili anche nel catalogo online del RISM; individuare e risolvere le duplicazioni presenti nelle basi dati SBN e RISM; implementare nell'OPAC SBN la ricerca e la visualizzazione dell'incipit musicale.

Anche in *Internet Culturale*⁶, la teca digitale delle biblioteche italiane, si possono consultare numerose risorse sonore tramite il *player* dedicato che consente di accedere direttamente alla risorsa sonora e ascoltare, per intero se fuori diritti, o per soli 30 secondi, le tracce fornite dai partner del portale. Attualmente si tratta di oltre 135.000 notizie bibliografiche che corrispondono a quasi 35.000 monografie e circa 100.000 spogli.

Le risorse digitali dell'ICBSA hanno dato origine a tre collezioni di *Internet culturale: Fondo storico cilindri fonografici, Fondo storico dischi e Fondo Compact Disc* dove sono consultabili tanto risorse musicali, quanto fonti orali (discorsi, brani teatrali e letterari, programmi radiofonici). Per quanto riguarda le fonti per la tradizione orale è anche significativa la presenza del *Fondo Majorano* che contiene canti della tradizione orale di Taranto e provincia, raccolti dall'etnografo Alfredo Majorano Nunziato⁷.

Musica, voci, discorsi e canzoni sono anche presenti nel portale tematico *14-18: Documenti e immagini della Grande Guerra*⁸, che raccoglie una documentazione quanto mai eterogenea sul primo conflitto mondiale, proveniente da 113 istituti di conservazione, per un totale di 645.000 risorse digitali, da anni in continua crescita.

Proprio a partire da queste risorse, nel novembre del 2016, in accordo con l'ICBSA e con Wikimedia Italia, è stata promossa una esperienza di riuso: l'*Editathon verdiano*, sulle opere di Giuseppe Verdi. In quella occasione, grazie alle tracce sonore, in versione integrale e libere dai diritti, concesse dall'ICBSA, gli esperti che hanno aderito e partecipato all'iniziativa hanno integrato, innalzandone il livello di autorevolezza, 19 voci esistenti in Wikipedia e ne hanno pubblicate 12 del tutto nuove.

Per consentire ad un pubblico, che ci auguriamo sempre più vasto, la consultazione integrata di tali risorse multidisciplinari presenti nelle varie basi dati gestite

⁵ https://www.iccu.sbn.it/export/sites/iccu/documenti/2016/Convenzione_IALM_ICCU_RISM.pdf.

⁶ <http://www.internetculturale.it/>.

⁷ <http://www.internetculturale.it/it/41/collezioni-digitali/26306/archivio-sonoro-alfredo-majorano>.

⁸ <http://www.14-18.it/>.

dall'ICCU (SBN, Internet culturale, il portale tematico 14-18, ma anche il sito delle mostre virtuali, MOVIO, il censimento dei manoscritti antichi e moderni *Manus OnLine* e quello delle edizioni italiane del XVI secolo EDIT 16), l'ICCU ha recentemente avviato un progetto, della durata di 24 mesi, che consentirà di navigare fra record di catalogo e risorse digitali a partire da un unico Portale delle biblioteche e degli istituti culturali italiani, una piattaforma di servizi informativi in grado di offrire al pubblico una nuova esperienza di ricerca basata sulla scoperta e la navigazione tra risorse di natura diversa ma riferibili a categorie di interesse come luoghi, persone, istituzioni⁹.

Per quanto riguarda le basi dati appositamente create per lo sviluppo della ricerca scientifica in ambito linguistico, glottologico, fonologico, l'esperienza dell'ICCU si riferisce in primo luogo alla partecipazione, negli ultimi dodici anni, a una linea di progettualità europea dedicata specificamente alle *e-infrastructures*, piattaforme digitali per la ricerca, che costituiscono oggi un fattore unificante contro la frammentarietà esistente e una grande opportunità di condivisione e reciproco arricchimento offerta a università ed enti di ricerca che operano in tutti i paesi dell'Unione.

L'ICCU, in particolare, ha partecipato al progetto PARTHENOS¹⁰, finanziato nell'ambito di Horizon 2020 e formato da un consorzio di 16 partner, che ha riunito le grandi infrastrutture europee nel settore del patrimonio culturale e delle *digital humanities*, come DARIAH, CLARIN, INRIA e importanti centri di ricerca come il DANS in Olanda e il PIN-Polo universitario della città di Prato, con l'obiettivo di armonizzare standard, modelli di gestione dei dati, protocolli di interoperabilità per aumentare la visibilità e la riusabilità dei dati della ricerca.

Le linee guida di PARTHENOS¹¹, recentemente pubblicate in inglese, sono indirizzate sia ai produttori di dati, e dunque ai ricercatori nell'ambito di discipline umanistiche come la storia, l'archeologia, gli studi linguistici e le scienze sociali, che alle infrastrutture di ricerca e agli istituti che gestiscono archivi e repository di dati.

Attraverso il paradigma FAIR (acronimo di *Findable, Accessible, Interoperable and Reusable*¹²) le raccomandazioni richiamano, in venti semplici punti, principi e metodi per migliorare l'accessibilità e il riuso dei dati¹³. Le linee guida sono state tradotte in italiano a cura dell'ICCU, che ne promuoverà in Italia la diffusione, ma anche l'adattamento alla realtà nazionale, perché è evidente che le soluzioni dei problemi di interoperabilità e di riusabilità non dipendono solo dalla tecnologia, ma riguardano le politiche nazionali, le licenze adottate, le restrizioni legali, gli aspetti etici collegati alla tutela della privacy e, non da ultimo, le figure professionali necessarie (su questo tema il progetto Parthenos è intervenuto offrendo una sezione di *training on-line* che presenta le migliori pratiche nella gestione dei dati di ricerca).

⁹ <https://www.iccu.sbn.it/it/eventi-novita/novita/Portale-delle-biblioteche-e-degli-istituti-culturali-italiani/>.

¹⁰ <http://www.parthenos-project.eu>.

¹¹ http://www.parthenos-project.eu/portal/policies_guidelines.

¹² FAIR, (2006). The FAIR Guiding Principles for scientific data management and stewardship, *Scientific Data*, 3.

¹³ <https://zenodo.org/record/2668479#.XOqn5CZ7m3A>.

Per favorire la conservazione, la descrizione, il riutilizzo di risorse orali raccolte in ambito di progetti di ricerca è necessario dunque prevedere fin dall'inizio l'uso di standard che garantiscano la qualità dei dati e un buon livello FAIR e puntare sulla completezza dei metadati, che devono accogliere sia le informazioni descrittive della risorsa, che gli elementi tecnici e gestionali, necessari anche per la conservazione a lungo termine dell'oggetto digitale.

Anche *Europeana*¹⁴, il portale del patrimonio culturale digitale della UE, ha definito un livello minimo di metadati (che può essere eventualmente arricchito con ulteriori informazioni), che consente a tutti gli aggregatori nazionali di inviare i dati. Questi – a loro volta – condividono con la rete nazionale degli istituti culturali gli standard per l'aggregazione dei metadati e la qualità degli oggetti digitali, oltre che le indicazioni sulle licenze da adottare.

Protocolli di scambio e standard descrittivi rappresentano dunque il presupposto per la costituzione di una rete tra associazioni, istituzioni, enti di ricerca, dedicate alla cura di dati trasversali a molte e diverse discipline.

Per quanto riguarda l'Italia, che vanta una collaborazione ultradecennale con *Europeana*, il ruolo chiave nel flusso di risorse digitali legate al patrimonio culturale italiano verso il portale europeo è svolto dall'aggregatore nazionale *CulturaItalia*¹⁵, gestito dall'ICCU, che – insieme a *Internet culturale* – costituisce il ponte che collega le istituzioni culturali del nostro paese con la grande infrastruttura europea per il patrimonio culturale digitale, con cui condivide standard e *policy* di pubblicazione dei dati.

Così è avvenuto per il progetto tematico *Europeana Sounds* (2014-2017), guidato dalla British Library con l'intento di facilitare l'accesso e migliorare la visibilità dell'enorme patrimonio sonoro conservato presso le istituzioni Europee, nell'ambito del quale l'ICCU ha aggregato e reso disponibili 83.000 risorse sonore fornite dall'ICBSA.

A parte i notevoli risultati quantitativi (520.000 risorse provenienti da 24 istituzioni di 12 paesi europei) che hanno fatto di *Europeana Sounds* uno dei progetti tematici di maggiore successo nella grande famiglia di *Europeana*, un importante contributo allo studio delle politiche comuni per migliorare l'accesso al patrimonio sonoro digitalizzato è anche rappresentato dal documento *Europeana Sounds Policy Recommendations for improving online access to audio and audio-related heritage in Europe*¹⁶ dove viene direttamente affrontata la delicata tematica del diritto d'autore, dal punto di vista etico, oltre che legale, che rappresenta una caratteristica peculiare delle fonti orali e richiede alla Commissione europea un impegno ulteriore per lo sviluppo di un bilanciamento tra la legislazione sulla protezione dei dati, la legge sulla privacy e la missione delle istituzioni del patrimonio culturale di rendere accessibili le proprie collezioni al pubblico.

¹⁴ <https://www.europeana.eu/portal/it>.

¹⁵ <http://www.culturaitalia.it/>.

¹⁶ https://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Sounds/Deliverables/europeanasonsounds-d3.4-policy-recommendations.pdf.

