

Il flusso di lavoro nei progetti di aggregazione di contenuti culturali digitali. Buone pratiche e controllo della qualità

Marzia Piccininno - ICCU

Valentina Vassallo - The Cyprus Institute - STARC

Uno degli obiettivi dei progetti di aggregazione di contenuti culturali digitali è quello di raccogliere grandi quantità di metadati forniti da molteplici istituzioni culturali, standardizzandone i formati e rendendoli consultabili da un unico punto di accesso. Ciò implica la necessità di stabilire una rigorosa procedura in grado di monitorare tutti i passaggi, dalla fornitura alla pubblicazione, e di definire buone pratiche e soluzioni per la sostenibilità a lungo termine dei contenuti aggregati. Il contenuto di questo contributo non è solo tecnico, ma anche procedurale poiché prende in considerazione tutte le strutture di *back-end* dell'aggregazione di contenuti culturali eterogenei. Sarà presentato lo sviluppo di un flusso di lavoro così come avvenuto in alcuni progetti di aggregazione per Europeana, il portale dei contenuti culturali degli Stati Membri, finanziati dalla Commissione europea (ATHENA, Linked Heritage).

1. Introduzione

È opinione errata, ma comune che in un aggregatore di risorse culturali digitali la parte tecnologica sia preponderante, il cuore di tutto il sistema, a scapito di altri fattori altrettanto importanti quali la gestione dei contenuti e la loro qualità.

Questo contributo nasce dall'esperienza delle autrici nel campo dei progetti europei di aggregazione di risorse culturali digitali, in particolare nella gestione del flusso di lavoro verso Europeana, la più grande biblioteca digitale culturale in Europa. Questa esperienza si è svolta nel 'Coordinamento dei contenuti' per i progetti europei ATHENA, Linked Heritage e, attualmente, AthenaPlus e ha fatto emergere l'importanza e la necessità di organizzare un rigoroso flusso di lavoro per la fornitura di contenuti ad Europeana. Un'istituzione culturale può contribuire a Europeana direttamente o tramite aggregatori, ma avrà sempre la necessità di un guida per aggregare i propri dati nel rispetto degli standard di Europeana. Ciò implica l'esigenza di un approccio standardizzato per tutte le fasi del flusso di lavoro, dalla mappatura dei diversi schemi di metadati (*mapping*), al rispetto dei requisiti minimi richiesti da Europeana, fino alla pubblicazione sul portale europeo.

La metodologia sviluppata nei tre progetti citati riguarda la gestione dei metadati – non degli oggetti digitali che rimangono sotto la responsabilità del fornitore di contenuti – e fa riferimento alle buone pratiche e ai vincoli che regolano il flusso di lavoro di ingestione, descrizione delle norme e del quadro strategico.

2. Aggregare contenuti per Europeana

Lo scopo dei progetti di aggregazione di contenuti culturali digitali è quello di rendere accessibili attraverso un unico punto di accesso risorse provenienti da istituzioni di diverso tipo (archivi, biblioteche, musei) che utilizzano differenti schemi per descrivere i propri contenuti digitali. Le iniziative citate in questo articolo, ATHENA, Linked Heritage e AthenaPlus, tutte coordinate dall'ICCU, hanno avuto e hanno come principale obiettivo quello di svolgere questo lavoro per Europeana, la biblioteca digitale europea.

Europeana¹ è un'iniziativa della Commissione Europea nata su impulso degli Stati Membri; attualmente dà accesso a oltre 27 milioni di metadati culturali pertinenti a risorse digitalizzate in tutta Europa (libri, dipinti, documenti d'archivio, video, manufatti archeologici, film, fotografie, ecc.). Il lancio nel 2008 di Europeana è stato un punto di svolta nel panorama della digitalizzazione del patrimonio culturale; fino ad allora infatti l'accesso *online* ai contenuti culturali era possibile attraverso banche dati o *digital library* gestite dalle singole istituzioni - e dunque non integrate tra loro - o, in pochi pionieristici casi, attraverso i portali nazionali della cultura che organizzavano il sapere su base multisettoriale (*cross-domain*)². Europeana invece, proponendosi come punto di accesso unico ai contenuti culturali europei di ogni settore, tipologia e provenienza, ha contribuito a dare un grande impulso alla nascita di aggregatori su base nazionale e regionale, ma anche tematici o di settore³.

“Un aggregatore nell'ambito di Europeana è un'organizzazione che raccoglie i metadati da un gruppo di fornitori di contenuti e lo trasmette ad Europeana. Gli aggregatori raccolgono il materiale dalle singole organizzazioni, standardizzano i formati e i metadati e convogliano questi ultimi in Europeana secondo le sue linee guida e le sue procedure. Gli aggregatori supportano anche i fornitori di contenuti attraverso attività di gestione e formazione”⁴.

¹ www.europeana.eu. Il portale è gestito dalla Europeana Foundation, ospitata presso la Biblioteca nazionale olandese con sede a L'Aia. Il progetto è finanziato in larga parte dalla Commissione europea; a carico degli Stati Membri restano i costi della digitalizzazione.

² Nel 2008 era già attivo da diverso tempo *Moteur Collections*, il portale delle risorse digitali culturali gestito dal Ministero francese: <<http://www.culture.fr/Ressources/Moteur-Collections>>; lo stesso anno fu lanciato CulturalItalia il portale italiano promosso dal Ministero dei beni e delle attività culturali e del turismo e gestito dall'Istituto centrale per il catalogo unico delle biblioteche italiane – ICCU: <<http://www.culturaitalia.it/>>.

³ A tale proposito cfr. su questa rivista, Piccininno 2009.

⁴ “*The Europeana Aggregator Handbook 2011*”, p. 8, disponibile all'indirizzo: <<http://pro.europeana.eu/documents/858566/858665/Aggregators+Handbook>>.

Un aggregatore mira a creare forti partenariati, agendo da ponte tra le singole istituzioni culturali e Europea, e impostando le regole che possono garantire un armonioso processo di aggregazione e la fornitura di contenuti di qualità. Questo è il ruolo dei progetti di aggregazione nell'ecosistema Europea⁵ e dunque di ATHENA, Linked Heritage e AthenaPlus⁵.

ATHENA (2008-2011)⁶ è stato uno dei progetti più riusciti di questo gruppo: ha consentito l'aggregazione di oltre 4 milioni di contenuti digitali di centinaia di musei europei e ha consolidato meccanismi, raccomandazioni e strumenti per aiutare i partner, provenienti da oltre 23 Paesi, nella fornitura di contenuti a Europea. Tali innovazioni sono state immediatamente adottate da altri progetti europei e da iniziative nazionali e locali di aggregazione. Linked Heritage (2011-2013)⁷ sta proseguendo e ampliando il lavoro di ATHENA in una prospettiva *cross-domain*, coinvolgendo altri tipi di istituzioni culturali, mentre AthenaPlus (2013-2015)⁸ sperimenterà il riuso creativo dei dati museali in corso di aggregazione per Europea. Tra gli strumenti lasciati in eredità dal progetto ATHENA a tutti gli altri aggregatori è opportuno segnalare MINT⁹, un servizio web, basato su software open source, sviluppato dalla *National Technical University* di Atene (NTUA), che consente l'aggregazione di contenuti per Europea in modo semplice ed efficace. Grazie a MINT l'utente viene guidato nel percorso di aggregazione dei propri metadati per Europea e può mappare, convalidare e inviare dati ad Europea in un unico ambiente web¹⁰.

Il cuore di MINT è LIDO (*Lightweight Information Describing Objects*)¹¹, un formato di *harvesting* basato sul modello concettuale CIDOC-CRM nato per mantenere intatta la ricchezza semantica dei metadati di ambito museale ma, come sta dimostrando il progetto Linked Heritage, sufficientemente flessibile per essere utilizzato anche in ambito archivistico e bibliotecario¹². Grazie a MINT e LIDO i fornitori di contenuti di ATHENA prima, e quelli di Linked Heritage e AthenaPlus ora, sono in grado di caricare i propri dati, mappare il proprio schema di metadati

⁵ <http://pro.europeana.eu/projects>.

⁶ <http://www.athenaeurope.org/>.

⁷ <http://www.linkedheritage.org/>.

⁸ www.athenaplus.eu. Cfr. su questa rivista, Natale 2012.

⁹ <http://mint.image.ece.ntua.gr/>.

¹⁰ MINT viene personalizzato in base alle differenti esigenze dei vari progetti di aggregazione; in questo testo si fa riferimento alla versione approntata per il progetto Linked Heritage.

¹¹ La rappresentazione di LIDO in XML con tutte le specifiche si trova all'indirizzo www.lido-schema.org. Cfr. Coburn *et alii*, 2010. LIDO è presto divenuto uno standard internazionale. È infatti utilizzato come schema per alcuni aggregatori nazionali come la Biblioteca digitale tedesca: <http://www.deutsche-digitale-bibliothek.de/> e – in prospettiva – la Biblioteca nazionale digitale della Finlandia (Autere *et alii*, 2011).

¹² Gli usi di LIDO sono documentati all'indirizzo: <http://network.icom.museum/cidoc/working-groups/data-harvesting-and-interchange/lido-community/use-of-lido/>. Nel corso di Linked Heritage sono state fatte diverse mappature tra LIDO e gli schemi della "famiglia" MARC.

su LIDO, verificare l'anteprima dei propri dati in Europeana, validarli e inviarli al portale europeo. MINT offre diversi vantaggi: trasforma i dati caricati e mappati in LIDO negli schemi in uso da parte di Europeana (ESE – *Europeana Semantic Element*¹³ allo stato attuale, EDM – *Europeana Data Model*¹⁴ nell'immediato futuro), e conserva i metadati forniti dalle varie istituzioni in un unico *repository* gestito da NTUA, garantendo la sostenibilità della fornitura dei contenuti oltre la fine dei progetti; inoltre consente la validazione dei contenuti attraverso l'*Europeana Content Checker*¹⁵, la conversione dei dati archiviati in ESE¹⁶ e la trasmissione al gruppo di ingestione dati di Europeana attraverso il protocollo di esportazione OAI-PMH¹⁷.

L'utilizzo di MINT consente di impostare un flusso di aggregazione di metadati standardizzato, monitorabile in tutte le sue fasi, e dunque facilmente replicabile in altri contesti simili.

2.1. Aggregare metadati: la parte tecnica

LIDO è un formato di *harvesting* basato sul modello di riferimento concettuale CIDOC-CRM¹⁸. Proviene dall'integrazione tra il CDWA Lite, lo schema XML elaborato dalla fondazione Getty per la descrizione degli oggetti culturali e dei loro "surrogati" digitali¹⁹, e museumdat²⁰, il formato di *harvesting* in uso nell'ambito museale, e si basa sullo standard britannico SPECTRUM²¹.

LIDO v.1.0²², integrato nella versione di MINT per Linked Heritage, è organizzato in 7 aree di cui 4 sono descrittive e 3 amministrative (tavola 1)²³.

¹³ <http://pro.europeana.eu/technical-requirements>.

¹⁴ <http://pro.europeana.eu/web/guest/edm-documentation>.

¹⁵ *Europeana Content Checker version 2 User Guide*:

<<http://pro.europeana.eu/documents/900548/ae5e78e8-ce78-424d-b360-5c01eddb3564>>.

¹⁶ ESE sta per *Europeana Semantic Element*, il modello di metadati utilizzato da Europeana. MINT supporta anche EDM (*Europeana Data Model*), il nuovo profilo di applicazione che sarà implementato a breve.

¹⁷ La trasmissione dei dati ad Europeana è tecnicamente gestita da NTUA, tuttavia Europeana ha chiesto ai fornitori di contenuti di verificare ulteriormente la qualità dei contenuti una volta pubblicati on-line, al fine di accertare eventuali problemi.

¹⁸ Il modello di riferimento concettuale CIDOC (CRM) fornisce definizioni e una struttura formale per descrivere i concetti e le relazioni implicite ed esplicite utilizzati nella documentazione del patrimonio culturale: <<http://www.cidoc-crm.org/>>.

¹⁹ http://www.getty.edu/research/publications/electronic_publications/cdwa/cdwalite.html.

²⁰ <http://www.museumdat.org/index.php?ln=en>.

²¹ <http://www.collectionslink.org.uk/spectrum>.

²² <http://www.lido-schema.org/schema/v1.0/lido-v1.0-specification.pdf>.

²³ Cfr. su questa rivista Stein 2010.

DESCRIPTIVE INFORMATION	ADMINISTRATIVE INFORMATION
<ul style="list-style-type: none"> - Object Classification <ul style="list-style-type: none"> • Object type (mandatory) • Classification - Object Identification <ul style="list-style-type: none"> • Title / Name (mandatory) • Inscriptions • Repository • Display / Edition • Description • Measurements - Event <ul style="list-style-type: none"> • Event ID • Event type, • Object's role in the event • Event name • Actors • Cultures involved • Date • Period • Places • Event method • Materials and techniques • Other objects present at the event • Related events • Description of the event - Relation <ul style="list-style-type: none"> • Subject Set • Related Objects 	<ul style="list-style-type: none"> - Rights Work <ul style="list-style-type: none"> • Rights type • Rights holder • Rights dates • Credit line - Record <ul style="list-style-type: none"> • Record ID (mandatory) • Record Type (mandatory) • Record Source (mandatory) • Record rights • Metadata references - Resource <ul style="list-style-type: none"> • Link • Resource ID • Relationship type • Resource type • Resource rights • View description • View type • View date • Resource source • Related resources • Resource metadata location

Tavola 1. La struttura di LIDO

La piattaforma MINT utilizza LIDO come schema intermedio tra gli schemi di metadati in uso presso le varie istituzioni culturali per descrivere le loro risorse digitali (da quelli maggiormente utilizzati come *Dublin Core* e MARC, a quelli proprietari) e ESE/EDM per la pubblicazione in Europeana.

La necessità di utilizzare LIDO come schema intermedio per armonizzare i metadati per Europeana nasce dall'esigenza espressa dai musei che parteciparono al progetto ATHENA di mantenere integra la ricchezza delle descrizioni degli oggetti museali; l'*Europeana Semantic Element* usato all'epoca, basandosi sul *Dublin Core*, era alquanto insoddisfacente per i curatori museali in quanto, una volta avvenuta la conversione dei metadati per Europeana, le informazioni a corredo delle loro risorse digitali risultavano fortemente impoverite.

LIDO invece ha consentito di mantenere integra la ricchezza semantica dei dati prodotti in ambito museale.

Le organizzazioni che effettuano la registrazione in MINT possono caricare i loro metadati in formato XML o CSV utilizzando la modalità di esportazione più congeniale (http, ftp, OAI-PMH) al fine di gestire, aggregare e pubblicare in Europea le loro collezioni.

MINT è facile da usare: infatti, gli utenti possono definire i loro *mapping* con LIDO attraverso un *editor* visuale per il linguaggio XSL. Il *mapping* viene eseguito tramite semplici operazioni di *drag-and-drop* o di input che poi vengono tradotti nel codice corrispondente (figura 1)²⁵.

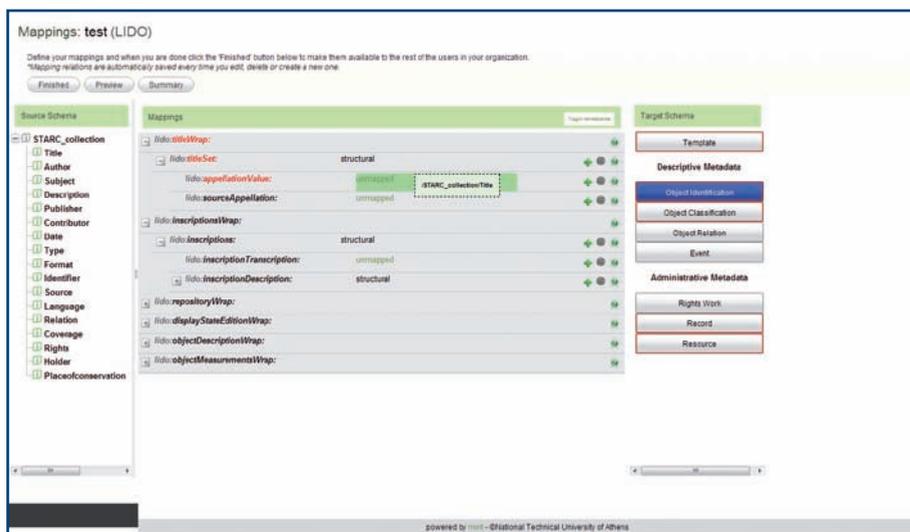


Figura 1. Un esempio della procedura di mapping da uno schema di metadati proprietario in LIDO, tramite lo strumento MINT.

L'editor di *mapping* mostra sia l'*input* che il *target* XSD, grazie all'interfaccia intuitiva che fornisce l'accesso e la navigazione, lo schema di *input* dei dati, la struttura, la documentazione e le restrizioni dello schema di output. Questi *mapping* possono essere applicati ai metadati caricati su MINT, ma anche modificati, scaricati e condivisi come modelli tra gli utenti della piattaforma.

Un'interfaccia di anteprima mostra agli utenti le fasi di aggregazione, tra cui l'XML caricato, l'XSLT della loro mappatura, il record trasformato nello schema di desti-

²⁵ Per l'utilizzo di MINT si può anche vedere il corso introduttivo in modalità *e-learning* elaborato dal Centro Bibliotecario di Ateneo dell'Università di Padova: <<http://linkedheritage.cab.unipd.it/training/LO-03/en/00.html>>.

nazione, le trasformazioni dallo schema di destinazione ad altri modelli di interesse (ad esempio ESE) e la rappresentazione html di ciascun record XML.

Sin da subito con il progetto ATHENA si è palesata la necessità di creare una metodologia di lavoro strutturata per regolare e monitorare tutte le fasi della procedura di *mapping* e caricamento dei metadati in MINT. Mappare il proprio schema di metadati su LIDO è l'attività più complessa e che dunque richiede più tempo. LIDO è infatti uno schema molto articolato in quanto preserva la ricchezza dei dati, ma proprio per questo il primo impatto può risultare complicato ed è preferibile che la procedura sia guidata in un primo momento da esperti.

2.2 Aggregare metadati: il flusso di lavoro

La definizione di un flusso di lavoro serrato in grado di organizzare la fornitura di contenuti a Europeana serve ad instradare le attività delle istituzioni culturali che forniscono contenuti e ricevere dei riscontri puntuali in merito alle loro difficoltà, i possibili miglioramenti della catena di lavoro ed eventuali problemi del software.

La metodologia di seguito illustrata permette di monitorare tutta la filiera di ingestione, dalla stima numerica dei contenuti che si mettono a disposizione del progetto fino alla pubblicazione in Europeana.

a. *Valutazione delle collezioni digitali*

Può sembrare banale, ma il primo passo da compiere per assicurare una fornitura di risorse di qualità a Europeana è quello di valutare le collezioni digitali che i fornitori di contenuti intendono mettere a disposizione dei progetti di aggregazione. L'esperienza del progetto ATHENA, e di Linked Heritage attualmente, dimostra che dalla scrittura della proposta progettuale al suo effettivo finanziamento possono accadere molte cose che influenzano la disponibilità della collezione digitale: problemi tecnici (reingegnerizzazione della piattaforma tecnologica per la gestione dei dati, metadati senza oggetti digitali collegati, ecc.), mancanza di autorizzazione per la pubblicazione *online*, cambiamento del funzionario responsabile, ecc. Questa operazione è facilmente effettuabile con un semplice questionario i cui risultati saranno registrati in un piano di aggregazione. È fondamentale chiedere di esplicitare la quantità di metadati e la quantità di oggetti digitali che i metadati descrivono poiché il rapporto 1:1 tra metadato e oggetto digitale non è sempre valido: istituzioni culturali differenti possono contare i risultati delle attività di digitalizzazione in modo diverso; Europeana inoltre aggrega (e dunque conta) solamente metadati, particolare che non è chiaro a tutte le istituzioni che intendono contribuirvi.

b. *Il piano di aggregazione*

La revisione dell'elenco dei contenuti dichiarati nella progetto consente di elaborare il piano di aggregazione. L'obiettivo di questo piano è di definire i tempi

dell'aggregazione, mantenere sotto controllo la fornitura di contenuti e monitorare i progressi.

c. *Dossier del fornitore*

A integrazione del piano di aggregazione, è opportuno realizzare un dossier specifico del fornitore di contenuti in cui tenere traccia di tutte le singole attività e delle variazioni rispetto al piano originale.

d. *Formazione*

La formazione in presenza sul *mapping* e sull'uso di MINT si è rivelata uno strumento prezioso di apprendimento e di condivisione dei problemi, che forniscono il giusto complemento alle linee guida e ai materiali didattici predisposti per rendere più fluido il flusso di lavoro. La formazione è poi resa costante dalla presenza di un servizio di *help-desk* (vedi *infra*).

2.3 Aggregare metadati: la raccolta dei feedback dei fornitori di dati

Di fondamentale importanza è la metodologia per la raccolta dei *feedback* da parte degli utenti per la risoluzione dei problemi e il miglioramento del processo di aggregazione. Gli strumenti a disposizione di questa attività sono diversi:

a. *MINT Data Report*

Monitorare il procedere delle attività di aggregazione dei vari fornitori di dati, nel caso si usi il servizio web MINT, è facilitato dalla presenza di una funzionalità chiamata *Data Report* che offre una rapida visione del numero complessivo

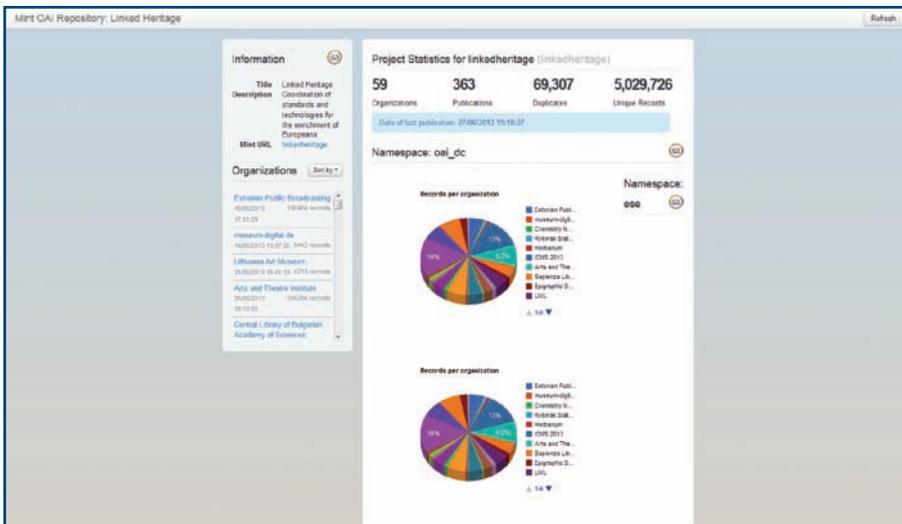


Figura 2. MINT OAI-PMH Repository Reports.

delle organizzazioni e degli utenti registrati, dei metadati che sono stati caricati e trasformati in LIDO e dei metadati pubblicati e quindi pronti per essere raccolti da Europeana.

Nell'ambito del progetto Linked Heritage è stata creato un ulteriore servizio a integrazione del *Data Report* che permette di tenere sotto controllo il passaggio dei metadati nel *repository* OAI-PMH, dopo la pubblicazione in MINT: il *MINT OAI-PMH Repository Reports* (figura 2). Questa nuova funzione consente innanzitutto di verificare la quantità di duplicati (i record, cioè, con la stessa URL e che non saranno trasferiti ad Europeana, al fine di garantire la presenza di risorse di qualità). Si tratta in pratica dell'interfaccia amministrativa "leggibile" del *repository* OAI-PMH.

b. Il servizio di *help-desk*

Non appena la fase di aggregazione di risorse è stata avviata, è ravvisabile la creazione di un servizio di *help-desk* per aiutare le istituzioni culturali a superare i loro problemi e a mantenere informati i coordinatori dei contenuti sul progresso delle attività.

Nonostante l'abbondante presenza di piattaforme di tipo *social*, si è ravvisato che la modalità più apprezzata per la condivisione dei problemi è la *mailing list*, in cui le persone interessate si registrano ed inviano le loro domande su vari argomenti: il *mapping*, il *software* (*bug* ed informazioni), le scadenze, la pubblicazione in Europeana. Un gruppo di esperti risponde in relazione al proprio campo di esperienza. Per facilitare la distribuzione delle risposte, ogni fornitore di contenuti deve specificare il tipo di problema nell'oggetto della e-mail: la persona responsabile dell'argomento risponde e facilita il lavoro durante la fase di ingestione. In particolare, nell'ambito di Linked Heritage sono stati considerati quattro argomenti: la procedura di *mapping*, il software, l'ingestione e il flusso di lavoro. Il servizio ha dimostrato di essere un supporto importante sia nel fornire soluzioni sia nel condividere problemi ed esperienze durante tutto il processo di aggregazione.

c. *Interviste periodiche*

È opportuno che chi coordina il flusso di lavoro effettui periodicamente un '*tour de table*' dei fornitori di contenuti tramite e-mail o telefonate per ottenere riscontri sul lavoro che stanno portando avanti. L'informazione viene comparata con quella fornita nel *MINT Data Report* e registrata nel dossier del fornitore che tiene traccia di ogni scambio di e-mail o telefonata o incontro al fine di tenere sotto controllo il flusso delle informazioni, specie in presenza di un consistente numero di istituti che partecipano alle attività di aggregazione.

d. *Riunioni plenarie*

“Storicamente, gli incontri faccia a faccia hanno svolto un ruolo importante in campo sociale, e soprattutto nella vita politica, delle società occidentali e non occidentali”²⁶. Questa citazione illustra l’importanza che le riunioni hanno nella società e nella vita organizzativa. Al giorno d’oggi, i mezzi di comunicazione elettronici sono frequentemente utilizzati, ma è stato notato che i processi di gruppo e i risultati che richiedono coordinazione, consenso, tempistiche, persuasione, sono meno efficacemente raggiunti utilizzando modalità di comunicazione mediata dal computer²⁷. Considerando queste circostanze, le comunicazioni faccia a faccia sono suscettibili di essere più efficaci rispetto a quelle mediate dal computer e aiutano a gestire le complessità, i dubbi o le incertezze nei compiti. D’altra parte, in condizioni di minore pressione o urgenza, quando le decisioni e le risposte sono relativamente meno importanti, le modalità di comunicazione mediate dal computer rappresentano una scelta più appropriata. Tuttavia, quando vi è la necessità di una comunicazione più “ricca” (cioè è richiesta una complessa interazione sociale, quando i compiti e le decisioni sono complesse e quando vi è la necessità di una risposta concreta), le comunicazioni faccia a faccia sono forse un canale più efficace. Naturalmente, una miscela appropriata di entrambe, incontri faccia a faccia e comunicazioni mediate da computer, sono la migliore combinazione²⁸.

Considerando gli studi socio-psicologici citati come termine di riferimento²⁹, l’esperienza diretta delle riunioni plenarie ci insegna l’importanza degli incontri faccia a faccia con i partner del progetto nel corso di riunioni plenarie. Queste occasioni permettono un *feedback* diretto e una comunicazione in tempo reale. Un altro vantaggio degli incontri faccia a faccia è che forniscono un vero e proprio *forum* per i membri al fine di ottenere e dare sostegno.

2.4 Aggregare metadati. il flusso di lavoro verso Europeana

Il flusso di lavoro precedentemente descritto illustra i passaggi interni ai progetti, mentre l’effettiva pubblicazione in Europeana avviene successivamente. Un volta pubblicati in MINT i metadati vengono trasferiti automaticamente sull’interfaccia OAI-PMH affinché lo staff tecnico di Europeana possa effettuare l’*harvesting*. In base agli accordi presi con Europeana, l’*harvesting* dei *dataset* non avviene mai

²⁶ Schwartzman 1989.

²⁷ Il tipo di mezzo di comunicazione rischia di pregiudicare i risultati ‘quando vi è la necessità di esprimere delle emozioni, quando i compiti richiedono coordinamento e la sincronizzazione tra le attività dei membri, quando si sta tentando di convincere gli altri, o si richiede il consenso su questioni che sono influenzate da atteggiamenti o valori dei membri del gruppo’ (Strauss, SG, McGrath, JE, 1994).

²⁸ Maznevski, Chudoba 2000.

²⁹ Arvey R.D., Why Face-to-Face Business Meetings Matter. Consultato il 4th July 2012.

<<http://www.iacconline.org/content/files/WhyFace-to-FaceBusinessMeetingsMatter.pdf>>.

automaticamente ma deve essere periodicamente segnalato tramite comunicazione scritta; una volta accolta la segnalazione, Europeana avvia la pubblicazione e un processo di controllo della qualità dei dati i cui risultati vengono trasmessi ai coordinatori dei contenuti dei vari aggregatori affinché esaminino con i fornitori eventuali modifiche da apportare. Il fine ultimo di questo processo è quello di pubblicare in Europeana dati quanto più standardizzati.

3. Le lezioni apprese

3.1 L'Europeana Data Exchange Agreement (DEA): ostacolo o opportunità?

Nel settembre 2011 Europeana ha rilasciato il *Data Exchange Agreement* (DEA), una licenza che disciplina i rapporti con le istituzioni culturali che forniscono i metadati³⁰ in sostituzione dei due precedenti distinti contratti per i fornitori di contenuti e per gli aggregatori. A differenza degli accordi precedenti, il DEA prevede che i metadati che descrivono le risorse digitali siano soggetti alla licenza Creative Commons CC0 che di fatto rappresenta una rinuncia a qualsiasi diritto sulle risorse cui è applicata, che ricadono dunque nel pubblico dominio³¹; questa disciplina non si applica alle anteprime degli oggetti digitali fornite a Europeana (i *thumbnail*) che invece mantengono i diritti gravanti sugli oggetti digitali di cui sono il surrogato. Questo cambiamento radicale ha un obiettivo ben preciso per Europeana, quello di poter rilasciare i metadati del portale (e dunque - almeno per ora - solo le parti descrittive, non le anteprime) come *Linked Open Data* in modo da favorirne il riuso anche a fini commerciali, e dare un impulso alle industrie creative. Questo approccio è ben spiegato nel documento intitolato *Europeana Public Domain Charter*³², il cui assunto principale è che la digitalizzazione di opere di pubblico dominio non deve creare nuovi diritti gravanti su di esse.

Il *Data Exchange Agreement* ha suscitato un grande clamore in tutta la comunità del progetto ATHENA, costituita in prevalenza da musei i quali erano preoccupati per l'eventuale utilizzo commerciale dei metadati, che avrebbe dovuto essere esplicitamente escluso. I metadati creati dai curatori museali contengono spesso un campo descrittivo dell'oggetto molto ricco, che costituisce di per sé un piccolo saggio su cui gravano diritti morali ed intellettuali.

La questione si è trascinata anche nel corso del progetto Linked Heritage, mettendo a rischio la partecipazione di alcune istituzioni culturali che non accettavano le condizioni imposte da Europeana, assolutamente non negoziabili.

Per ovviare questo problema una *task-force* di partner di Linked Heritage ha elaborato una strategia che propone ai fornitori di contenuti tre opzioni per la pubblicazione dei metadati, sempre nel rispetto del DEA:

³⁰ <http://pro.europeana.eu/web/guest/data-exchange-agreement>.

³¹ <http://wiki.creativecommons.org/CC0>.

³² Edito nell'aprile 2010: <http://pro.europeana.eu/c/document_library/get_file?uuid=d542819d-d169-4240-9247-f96749113eaa&groupId=10602>.

a. *Set minimo*

Il fornitore di contenuti trasmette a Europeana solamente i campi obbligatori per ciascun metadato, lasciando che esso sia fruito nella sua pienezza solamente sul proprio sito. In tal modo solamente una minima parte dell'informazione culturale viene rilasciata come di pubblico dominio. Questa è stata l'opzione scelta dalle istituzioni più fortemente avverse al *Data Exchange Agreement*.

b. *Set intermedio*

Con questa opzione tutti i campi dei metadati sono trasmessi a Europeana sotto licenza CC0, ad eccezione degli elementi di LIDO che si traducono in *dc:description* in ESE. Ciò significa che nessuna descrizione dell'oggetto sarà resa visibile in Europeana, ma solamente sul sito del fornitore.

c. *Set completo*

I metadati sono pubblicati in Europeana in tutta la loro articolazione; su tutti i campi dei metadati grava la licenza CC0.

Queste tre opzioni hanno reso necessario l'aggiornamento della piattaforma di aggregazione MINT, tramite l'installazione di un filtro di opzione per selezionare quella preferita durante il processo di aggregazione (figura 3):

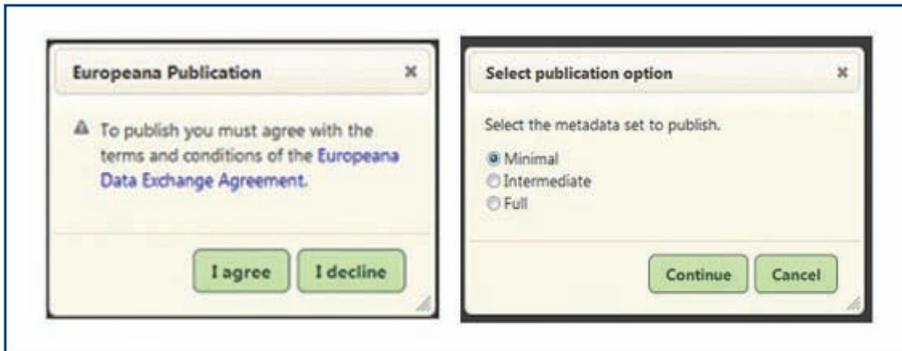


Figura 3. Selezione del filtro di opzione dei metadati in MINT

Una volta illustrati i dettagli del *Data Exchange Agreement* alle istituzioni culturali e offerte le tre opzioni per la fornitura di contenuti, è stato somministrato nell'ambito di *Linked Heritage* un sondaggio online³³ per analizzare lo stato della firma della licenza³⁴.

³³ I risultati hanno evidenziato che la maggior parte dei fornitori di contenuti di *Linked Heritage* ha firmato il DEA.

³⁴ Se da un lato *Linked Heritage* sta mettendo in luce tutte le possibili implicazioni che il DEA comporta, ciò non significa che il progetto incoraggi le istituzioni culturali a fornire a Europeana metadati con poche informazioni; ai partner, che non si sentono a proprio agio nel fornire set di

Questo approccio è altamente raccomandato perché il confronto tra istituzioni ha favorito la dissipazione dei dubbi sulla DEA e ha favorito la firma dell'accordo. Infatti, la possibilità di veder pubblicate le proprie risorse in Europeana senza rilasciarle totalmente nel pubblico dominio ha senza dubbio ammorbidito la posizione delle istituzioni culturali più avverse all'accordo. È superfluo sottolineare che fornire a Europeana metadati il più ricchi possibile è un obiettivo di qualità che dovrebbe essere perseguito da tutte le istituzioni che intendono prendervi parte.

3.2 Il controllo della qualità dei metadati

Un altro aspetto importante è il controllo della qualità dei contenuti aggregati ed è per questo che si è ritenuto utile mettere in atto sin dal progetto ATHENA un serie di strumenti affinché i dati forniti ad Europeana rispondessero a criteri tecnici e semantici di qualità.

Partendo dal presupposto che ogni istituzione culturale che produce risorse digitali ne è la sola responsabile del loro aggiornamento e della loro completezza, è opportuno che queste caratteristiche rimangano tali nei vari passaggi verso Europeana e che siano rispondenti ai parametri tecnici che essa richiede.

MINT è predisposto per segnalare nel corso delle mappatura e del caricamento dei metadati ogni deviazione dai requisiti di Europeana; gli utenti di MINT possono infatti controllare i loro contenuti attraverso la visualizzazione dell'*output* (l'xml mappato) e l'anteprima del risultato (come sarà, una volta pubblicato in Europeana), verificare se i metadati non rispettano i campi obbligatori richiesti da Europeana o se l'XML non risulta valido e la sua pubblicazione viene impedita. Questa azione viene ulteriormente effettuata a campione anche da chi coordina l'intero flusso di aggregazione dei partner verso il portale europeo.

Nel corso del progetto Linked Heritage sono stati resi più stringenti i parametri di controllo del risultato finale prima della pubblicazione dei dati grazie all'implementazione di una lista di controllo dei campi dei metadati durante il *mapping*; allo stesso tempo i fornitori di contenuti sono forzati a controllare un campione di dieci record prima di poter procedere alla pubblicazione.

4. Conclusioni

Il flusso di lavoro per guidare le istituzioni culturali nel loro processo di aggregazione di risorse per Europeana è stato realizzato e messo in atto nel corso dei progetti citati. Si è notato che la metodologia per la raccolta di *feedback* da parte dei fornitori di contenuti e per monitorare/coordinare le loro attività deve essere attivata sin dall'inizio del progetto.

metadati completi, viene semplicemente presentata la possibilità di scegliere il grado di granularità delle informazioni. Al fine di raggiungere obiettivi significativi, Linked Heritage si impegna a trovare soluzioni flessibili; in questo modo nulla viene aggiunto o rimosso dagli obblighi originali del progetto o dei suoi partner come stabilito nel Grant Agreement.

Il piano di aggregazione deve essere continuamente aggiornato per essere un efficace strumento di lavoro in grado di verificare l'effettiva disponibilità dei contenuti e monitorare l'avanzamento progetto dei lavori. La creazione di una comunità di utenti che persegue un unico obiettivo – veder pubblicati i propri dati in Europeana – è inoltre un elemento chiave per creare la coesione che aiuta a superare i problemi tecnici e procedurali di aggregazione.

La necessità di vedere pubblicati i propri dati in Europeana per un lungo periodo viene garantita dall'uso di uno strumento come MINT che, oltre a fornire al portale europeo metadati di qualità e standardizzati, ne consente la conservazione in un *repository* al quale i partner possono accedere per aggiornamenti anche dopo la fine del loro progetto di aggregazione.

Negli ultimi anni sono stati pubblicati alcuni studi su diverse problematiche relative ai metadati, ma solo di recente l'importanza della qualità dei metadati è diventata un importante argomento critico. Completezza, accuratezza e coerenza sono i criteri più comunemente usati nella valutazione della qualità dei metadati e per questo motivo è necessario realizzare un miglioramento dei sistemi per il controllo della qualità di tali elementi. Linked Heritage ha contribuito alla definizione di sistemi di controllo manuali, automatici e semiautomatici che garantiscono la presenza in Europeana solamente di dati di alto standard qualitativo.

Per riassumere, l'integrazione degli strumenti e delle procedure elaborate e descritte nel presente contributo hanno benefici sia per Europeana che per ogni iniziativa di aggregazione di contenuti digitali. Esse aiutano le istituzioni culturali che forniscono contenuti in ogni fase del loro lavoro e in parallelo permettono di monitorare i progressi delle attività, garantendo una migliore qualità contenuti delle risorse digitali fornite alla comunità.

L'ultima consultazione dei siti web è avvenuta nel mese di giugno 2013.

RIFERIMENTI BIBLIOGRAFICI

Autere R., Vakkari M., *Towards Cross-Organizational Interoperability: The LIDO XML Schema as a National Level Integration Tool for the National Digital Library of Finland*. Proceedings of the 15th international conference on Theory and practice of Digital Libraries: research and advanced technology for Digital Libraries, Springer-Verlag Berlin, Heidelberg 2011. ISBN: 978-3-642-24468-1.

Coburn E., Light R., McKenna G., Stein R., Vitzthum A., *LIDO - Lightweight Information Describing Objects Version 1.0*, 2010.

Maznevski M. L., Chudoba K. M., *Bridging space over time: Global virtual team dynamics and effectiveness*. Organizational Science, 1, 473-492, 2000.

McKenna G., Rohde-Enslin S., Stein R., *Lightweight Information Describing Objects (LIDO: the international harvesting standard for museums*, ATHENA project, 2010.

Natale M. T. (2012), *ATHENA: un ponte tra i musei europei ed Europeana*, in Digitalia, n. 2, pp. 58-61. <<http://digitalia.sbn.it/riviste/index.php/digitalia/article/view/238>>.

Park J-R, Tosaka Y., *Metadata Quality Control in Digital Repositories and Collections: Criteria, Semantics, and Mechanisms*. Cataloguing & Classification Quarterly. Volume 48, Issue 8, 2010.

Piccininno M., *Indagine sugli aggregatori europei di contenuti digitali*, Digitalia, n. 2, p. 171-185, 2009. <<http://digitalia.sbn.it/upload/documenti/Digit%202-2009.pdf?l=it>>.

Schwartzman H. B., (1989). *The Meeting: Gatherings in Organizations and Communities*. New York: Plenum Press.

Stein R., *Delivering content to Europeana in practice: the Athena harvesting format LIDO*, Digitalia, 5 (2010), n. 2, p. 157-161. <<http://digitalia.sbn.it/upload/documenti/DIGITA%202-2010.pdf?l=it>>.

Stein R., *Sharing Museum Information: Theory or Practice? A European Experience*. CIDOC 2010. ICOM General Conference Shanghai, China.

Strauss S.G., McGrath J. E. (1994). *Does the medium matter? The interaction of task type and technology on group performance and member reactions*. Journal of Applied Psychology, 79, 87-97.

Vassallo V., Piccininno M., *Aggregating Content for Europeana: a Workflow to Support Content Providers*, in Theory and Practice of Digital Libraries, TPD2012, 23-27 September 2012, Paphos, Cyprus, "Theory and Practice of Digital Libraries 2012", G. Buchanan, E. Rasmussen and F. Loizides (eds.), Springer.