



Progetto Far Conoscere WP3

Projet Faire Connaître WP3

PITEM PACE

Report di Raccomandazioni Tecniche

Regione Piemonte

Soggetti attuatori

Istituto di Informatica e Telematica - CNR

CSI Piemonte

Sommario

[1. Introduzione](#)

[2. Tipologie di risorse da digitalizzare](#)

[3. Raccomandazioni tecniche per la digitalizzazione](#)

[3.1 Immagini](#)

[3.2 Audio e Video](#)

[4. Raccomandazioni tecniche per la catalogazione](#)

[5. Interoperabilità](#)

[5.1 Tracciato di interscambio](#)

[5.2 Regole di nomenclatura dei file](#)

[5.3 Modalità di interscambio](#)

[5.4 Multilinguismo](#)

[5.5 Politiche di aggiornamento dei dati](#)

[6. Raccomandazioni tecniche per la valorizzazione e la promozione](#)

[Bibliografia](#)

Glossario

Risorsa Originale

Risorsa culturale così com'è descritta nel sistema nativo, come ad esempio un quadro.

Copia Digitale

Risultato di un processo di digitalizzazione di una risorsa originale. Ad una risorsa originale possono corrispondere più copie digitali, come ad esempio diverse foto aventi come oggetto lo stesso quadro.

Metadato

Attributo che specifica una proprietà della risorsa originale o della copia digitale (ad esempio il titolo di una risorsa originale o il formato della copia digitale).

Record

L'insieme di tutti i metadati relativi a una risorsa originale e alle copie digitali associate.

Tracciato di interoperabilità

Schema delle collezioni di record fornite all'aggregatore.

Formato del tracciato di interoperabilità

Uno tra i seguenti: xml, json, csv, ...

1. Introduzione

Questo documento descrive le raccomandazioni tecniche di cui ciascun partner dovrebbe tener conto per la digitalizzazione, la catalogazione e la valorizzazione delle proprie risorse culturali e l'interoperabilità dei propri sistemi di gestione. Le raccomandazioni proposte sono da considerarsi orientative, in quanto dipendenti dalle risorse culturali da digitalizzare, da valutare caso per caso a seconda del progetto considerato.

Dal documento di analisi tecnica emerge che i sistemi di gestione messi a disposizione dei partner come data provider per l'aggregatore sono i seguenti:

- Biblioteca Digitale Ligure (BDL) (Regione Liguria)
- Giornali del Piemonte (Regione Piemonte)
- Mèmora (Regione Piemonte)
- Catalogo dei Beni Culturali (Regione Valle d'Aosta)
- Syracuse (Ville de Nice).

Le risorse che verranno messe a disposizione per la fruizione attraverso l'aggregatore dovranno rispettare i seguenti requisiti:

1. ad ogni risorsa descritta deve essere associata almeno una copia digitale;
2. è possibile accedere alla copia digitale attraverso un URI;
3. la risorsa è correlata ad almeno uno dei temi del progetto PITEM Pa.CE;
4. la risorsa è rilasciata secondo una licenza che ne permette l'utilizzo;
5. la responsabilità della risorsa fa capo all'ente titolare del dato, che è quello che detiene il bene.

2. Tipologie di risorse da digitalizzare

Per analizzare i requisiti per le nuove digitalizzazioni è stato chiesto ai partner di indicare nel questionario non solo il numero delle nuove risorse che intendono digitalizzare, ma anche di che tipologia saranno.

Dai dati ricevuti si desume che i tipi dichiarati sono i seguenti: foto, mappe, disegni, lastre, diapositive, audio, video, giornali, libri, periodici. Si può quindi ricavare che i formati per i quali saranno necessarie delle linee guida per la digitalizzazione di nuove risorse saranno essenzialmente quattro: immagine, audio, video, testo.

Per *immagine* si intende di solito un artefatto bidimensionale che rappresenta un soggetto. Nel caso di una foto di un oggetto d'arte (ad esempio una statua) la foto è la rappresentazione bidimensionale di un soggetto tridimensionale. In questa sede intendiamo sia immagini originate dalla cattura tramite strumenti ottici (macchine fotografiche), quindi fotografie o diapositive, sia qualsiasi rappresentazione bidimensionale come una mappa, un grafico, un disegno, un dipinto, un pannello.

Le immagini in movimento (con o senza audio) rientrano nel formato *video*, mentre per *audio* intendiamo registrazioni di suoni senza immagini.

Come *testo* si considera un oggetto che possa essere letto, quindi per i nostri scopi rientrano nel formato testuale documenti, libri, giornali, periodici.

3. Raccomandazioni tecniche per la digitalizzazione

Nell'ambito del patrimonio culturale esistono varie tipologie di risorse culturali, ciascuna caratterizzata dal proprio formato, dal materiale e dal tipo di documento. Per poter effettuare una digitalizzazione occorre assumere che il documento sia piano e che la presenza di eventuali rugosità o pieghe non debba compromettere il processo digitalizzazione.

A prescindere dalla tipologia delle risorse considerate, generalmente esistono almeno le seguenti tre memorizzazioni su disco della stessa risorsa:

- **master** (a scopo di conservazione) - la rappresentazione deve essere effettuata in modo da garantire una qualità molto alta della risorsa; una risorsa master deve mantenere il più possibile tutta l'informazione in essa contenuta. Questo tipo di memorizzazione non è messa a condivisione con l'aggregatore, pertanto ogni partner può utilizzare le proprie linee guida per la memorizzazione master. Tuttavia, per completezza, si propongono delle raccomandazioni tecniche anche per questo tipo di memorizzazione;
- **dissemination** (destinata alla fruizione su browser Web) - la rappresentazione deve garantire un caricamento rapido da parte del browser, ma non deve essere troppo scadente. Pertanto, in questo caso si suggerisce una digitalizzazione che garantisca una media qualità della risorsa.
- **thumbnail/sample** - rappresentazione a bassa risoluzione, che permette di avere un'anteprima della risorsa da inserire ad esempio in una pagina contenente il risultato di una ricerca effettuata dall'utente.

Dal documento *Inventario dei siti Web* emerge che i modelli di diffusione maggiormente utilizzati dalle Digital Libraries for Cultural Heritage (DLCH) presentano i risultati utilizzando soltanto i livelli thumbnail e dissemination, ma non il livello master. Generalmente il livello thumbnail è prodotto automaticamente dalla DLCH a partire dall'immagine di dissemination. Pertanto, ai fini dell'interoperabilità con l'aggregatore, in teoria sarebbe sufficiente definire soltanto le linee guida di digitalizzazione per l'immagine di dissemination, lasciando liberi i partner di effettuare la digitalizzazione delle immagini master secondo le proprie esigenze. Tuttavia, per completezza, il presente documento fornisce anche delle linee guida per la digitalizzazione di immagini master e thumbnails.

Ogni tipologia di digitalizzazione è caratterizzata da requisiti minimi di digitalizzazione. Si suggerisce ai partner di rispettare le linee guida indicate in questo documento, al fine di garantire una certa omogeneità nella presentazione dei risultati sull'aggregatore.

3.1 Immagini

Il processo di digitalizzazione di un'immagine è una trasformazione dell'immagine continua (analogica) in una matrice di numeri interi. Tale processo si effettua in due successive fasi: il campionamento e la quantizzazione. Il campionamento consiste nel suddividere l'immagine in un reticolo di punti (campioni) detti pixel. A parità di grandezza dell'immagine, il numero di pixel

fornisce indicazioni sulla risoluzione dell'immagine, cioè sulla frequenza di campionamento spaziale applicata. La quantizzazione approssima il valore dei campioni ad un certo numero prefissato di valori. Generalmente la quantizzazione consiste nell'uso di una palette di colori per la rappresentazione dei valori assunti dai campioni:

- le immagini in bianco e nero sono quantizzate con 1 bit/pixel
- le immagini in grigio sono quantizzate tipicamente su 256 livelli (1 byte/pixel)
- le immagini a colori si scompongono in più canali (RGB, CMYB, etc.) e ciascun canale è quantizzato separatamente (raggiungendo 3/4 byte per pixel).

Durante il processo di digitalizzazione di immagini, vi sono alcuni errori tipici da evitare:

- immagine verticale acquisita nel modo paesaggio anziché ritratto (o viceversa)
- rotazione dell'originale
- immagine ribaltata
- problemi di luce con lo scanner
- immagine tagliata
- sporcizia nel ripiano dello scanner¹.

Come sintesi del documento di analisi tecnica emerge quanto segue:

Per quanto riguarda la digitalizzazione, tutti i partner effettuano digitalizzazioni ad una risoluzione minima delle immagini master (300 dpi) e per le scale di colore: 24 bit per le immagini a colore e 8 bit per le immagini in bianco e nero. Per le immagini di disseminazione (consultazione sul Web), non è possibile definire un valore minimo di risoluzione, perché vi sono diverse tipologie di immagini gestite, ognuna con una risoluzione diversa. Relativamente alla ricerca testo, i Giornali del Piemonte utilizzano il PDF con OCR integrato.

Master

Obiettivo dell'immagine master è mantenere tutta l'informazione, senza alcuna perdita. Per identificare la corretta risoluzione dell'immagine, si suggerisce l'utilizzo di tecniche di interpolazione, come specificato di seguito. L'interpolazione è il processo che partendo dai pixel reali presenti nell'immagine stima i pixel non conosciuti. Esistono diverse tecniche di interpolazione², che i diversi programmi per la digitalizzazione mettono a disposizione nell'elaborazione di immagini acquisite. Al fine di trovare la risoluzione ottimale per la digitalizzazione di un'immagine, in prima battuta, si prova ad effettuare due digitalizzazioni della stessa immagine, rispettivamente ad una risoluzione $R1 = X$ e $R2 = X + \square$. Successivamente si applica all'immagine con risoluzione $R1$ una tecnica di interpolazione in modo da ottenere la risoluzione $R2^*$. Si confrontano poi le immagini con risoluzioni $R2$ e $R2^*$. Se non si notano differenze tra le due immagini, allora la risoluzione ottimale per la digitalizzazione è $R1$, se si notano delle differenze, si procede ad un'ulteriore digitalizzazione dell'immagine ad una risoluzione $R3 = R2 + \square$, si effettua l'interpolazione di $R2$ in modo da ottenere $R3^*$ e così via fino ad ottenere la risoluzione ottimale.

Si suggerisce di digitalizzare immagini ad alta risoluzione in formato TIFF compresso o non compresso, con profondità delle riprese a colori (24 bit) o in toni di grigio (8 bit) con una risoluzione

¹ Fonte:

<https://www.sabap-abruzzo.beniculturali.it/wp-content/uploads/Normativa-per-lacquisizione-digitale-delle-immagini-fotografiche.pdf>

² <http://www.di.univr.it/documenti/OccorrenzaIns/matdid/matdid358544.pdf>

finale dipendente dalla tipologia del materiale. Generalmente, la risoluzione finale è intorno ai 300 dpi.

Dissemination

In questo caso l'obiettivo è la fruizione sul Web. L'immagine di dissemination si può ottenere effettuando un sottocampionamento dell'immagine master, attraverso appositi software di manipolazione di immagini, al fine di ottenere una risoluzione minore, utilizzabile per la fruizione sul Web.

In generale, la risoluzione ottimale per le immagini di dissemination dipende dal tipo di dispositivo utilizzato dall'utente finale. Le Figure 1, 2 e 3 mostrano le risoluzioni dello schermo maggiormente diffuse nel 2020, per tipologia di dispositivo (desktop, tablet e smartphone)³. Si suggerisce di utilizzare, per l'immagine di dissemination, la risoluzione maggiormente diffusa per pc desktop (1366 px x 768 px). Le altre risoluzioni possono essere ottenute come sottocampionamento della risoluzione proposta.

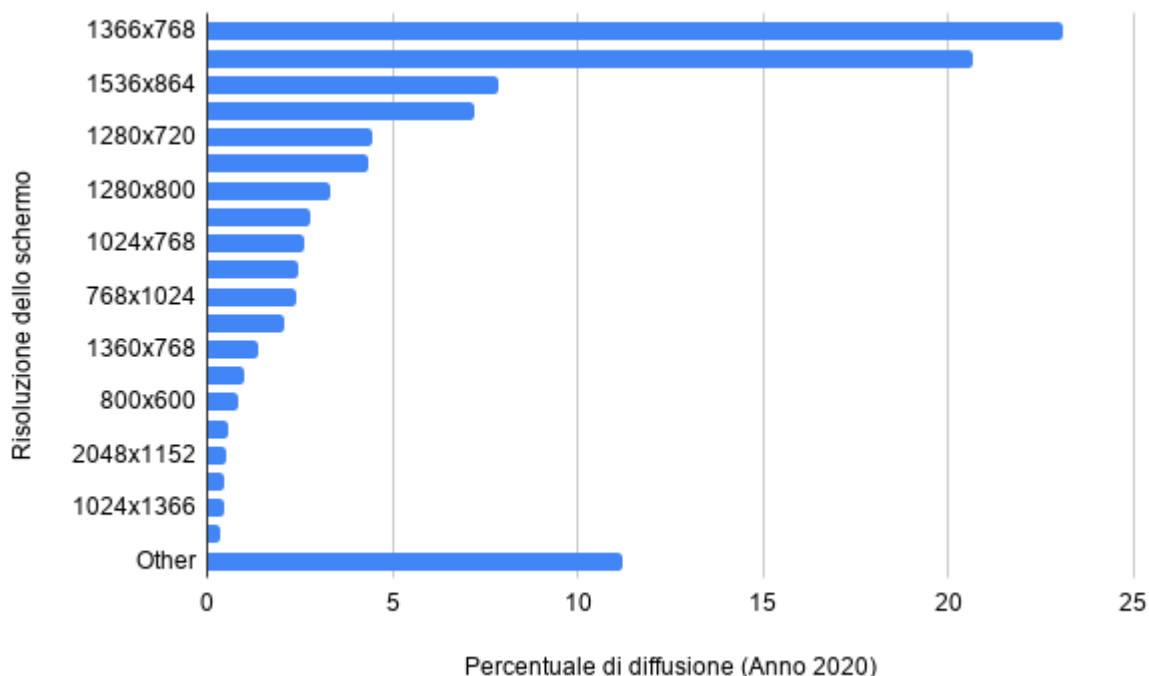


Fig. 1 Percentuale di diffusione delle varie risoluzioni dello schermo relativamente all'anno 2020 per PC desktop.

³ Fonte Stat Counter Global Stats: <https://gs.statcounter.com/screen-resolution-stats/>

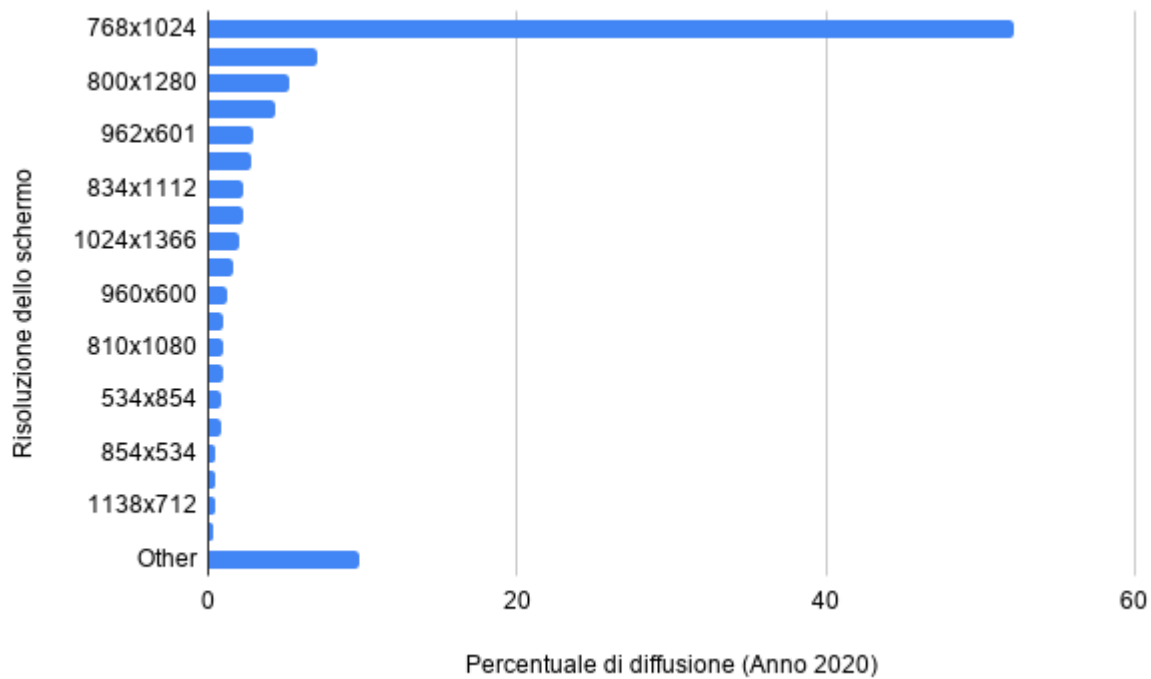


Fig. 2 Percentuale di diffusione delle varie risoluzioni dello schermo relativamente all'anno 2020 per tablet.

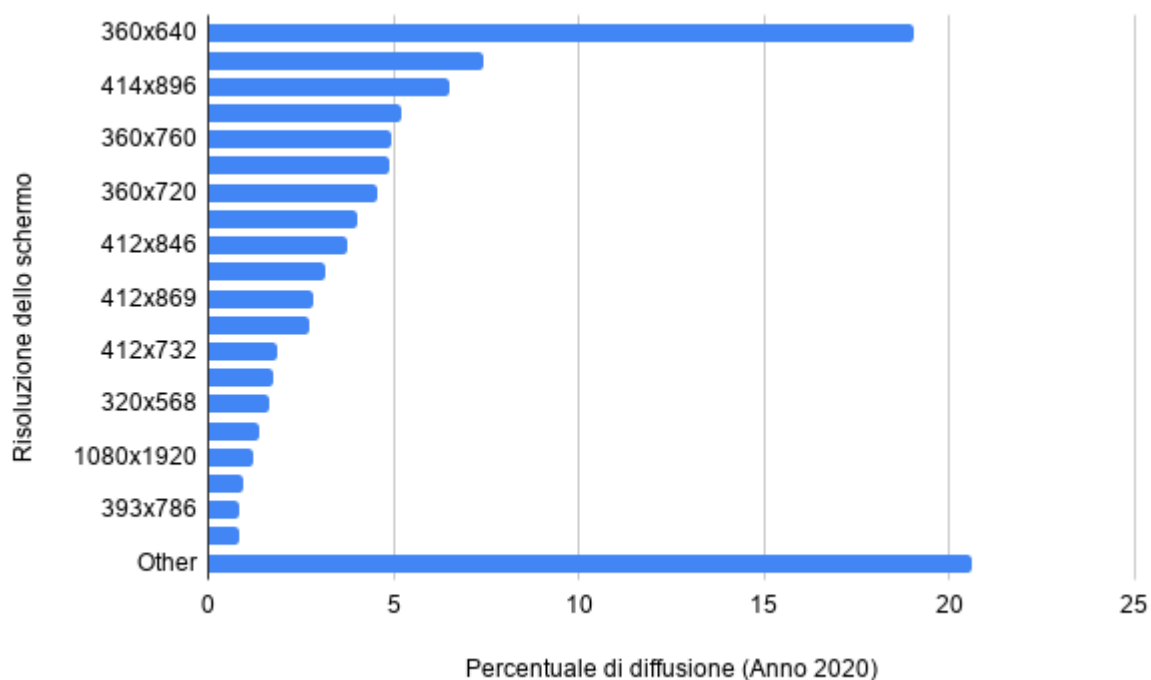


Fig. 3 Percentuale di diffusione delle varie risoluzioni dello schermo relativamente all'anno 2020 per smartphone.

Thumbnail

Digitalizzazione a bassa risoluzione, contenente un'anteprima dell'immagine o di un particolare che permetta all'occhio umano di capirne il contenuto. L'aggregatore estrarrà l'immagine

Thumbnail a partire dall'immagine di dissemination. Qualora tale immagine non fosse disponibile, l'aggregatore provvederà ad inserire un'immagine Thumbnail di default o un'immagine segnaposto.

3.2 Audio e Video

Anche per le risorse audiovisive vengono in genere create una copia master per la conservazione e delle copie compresse per la consultazione.

L'ICBSA⁴ (Istituto Centrale per i Beni Sonori e Audiovisivi) ha pubblicato le proprie linee guida per la digitalizzazione di documenti sonori e video [1]. La trascrizione in digitale è effettuata adottando le indicazioni dell'International Association of Sound and Audiovisual Archives (IASA)⁵, in conformità allo standard BWF prodotto dalla European Broadcasting Union (EBU)⁶ in collaborazione con la Audio Engineering Society (AES)⁷, accogliendo le indicazioni della risoluzione dell'Unesco sulla conservazione del digitale e le indicazioni pervenute dalle attività del progetto Presto (Preservation Technology for European Archives), un progetto relativo alla preservazione d'archivi analogici audio, video e film e guidato da BBC, RAI e Institut National de l'Audiovisuel con ACS e ITC/IRST come partner tecnici per la parte audio.

Viene raccomandato, per le copie master in alta qualità, l'uso dello standard europeo Broadcast Wave Format (BWF), un'estensione del formato WAV che consente di aggiungere nell'header del file alcuni metadati, senza richiedere un player specifico per la riproduzione.

Il campionamento del materiale analogico avviene a 48 kHz 24 bit stereo nei casi normali, e a 96 kHz 24 bit nei casi in cui il materiale debba essere sottoposto a restauro. Il materiale digitale nativo (ad esempio CD) viene trascritto mantenendo il campionamento originale.

Le copie compresse (per ridurre lo spazio di memorizzazione e la banda di trasmissione necessari) servono per la consultazione del materiale all'interno dell'Istituto o via Internet. Il bit rate e l'algoritmo di codifica possono variare in base agli avanzamenti tecnologici, ma dei valori utilizzabili sono 256 kbit/s con codifica MPEG 1 Layer 3 per la copia ad alta qualità, e 32 kbit/s con codifica MPEG 1 Layer 3 per quella qualità inferiore.

4. Raccomandazioni tecniche per la catalogazione

Come sintesi del documento di analisi tecnica emerge quanto segue:

Per quanto riguarda la catalogazione, i vari partner utilizzano standard e linee guida differenti: per le schede di catalogo, Biblioteca Digitale Ligure e SBNWeb utilizzano ICCU, mentre Mèmora e Catalogo dei Beni Culturali della Valle d'Aosta utilizzano ICCD. La Ville de Nice utilizza UNIMARC. Non c'è unanimità neanche sugli standard di catalogazione utilizzati: soltanto Mèmora e la Biblioteca Digitale Ligure utilizzano il Dublin Core.

Ai fini dell'interoperabilità tra i vari sistemi di gestione, non è necessario che le risorse digitali siano catalogate in modo uniforme e che tutti i sistemi utilizzino lo stesso standard di catalogazione o le

⁴ <http://www.icbsa.it/>

⁵ <https://www.iasa-web.org/>

⁶ <https://www.ebu.ch/home>

⁷ <http://www.aes.org/>

stesse linee guida. Ciò che conta è che tutti i sistemi di gestione utilizzino lo stesso tracciato di interoperabilità (descritto nella sezione seguente). Pertanto, per la catalogazione delle risorse, si suggerisce di utilizzare uno degli standard descritti nel documento di Analisi Tecnica.

5. Interoperabilità

5.1 Tracciato di interscambio

In questo capitolo viene descritto il tracciato di interoperabilità che si propone di utilizzare per la condivisione delle risorse che i sistemi dei partner identificheranno per la condivisione e la fruibilità attraverso l'aggregatore.

La proposta è volta ad uno standard che sia nello stesso tempo flessibile, in grado di rappresentare oggetti di diversa natura, riconosciuto come standard internazionale e che consenta l'esportazione dei dati in maniera semplice.

Lo standard che risponde a questi requisiti è il Dublin Core⁸, strutturato in due forme: il **simple** Dublin Core e il **qualified** Dublin Core. La prima forma è costituita da 15 elementi base, la seconda forma amplia con nuovi *tag* gli elementi base al fine di una descrizione più dettagliata del prodotto.

Le caratteristiche principali dello standard identificato sono:

- la **semplicità** di utilizzo in quanto il linguaggio è comprensibile sia da specialisti sia da neofiti;
- l'**estendibilità** che permette di aggiungere ai quindici elementi base nuovi elementi;
- l'utilizzo di una **semantica chiara e universalmente accettata** poiché creato da esperti di varie nazioni e orientato a garantire l'**interoperabilità** con altri profili di metadati.

La logica dello standard Dublin Core può essere così sintetizzata:

- un record DC è costituito da una o più proprietà, rappresentata dall'elemento, e dai valori associati
- ogni proprietà è un attributo della risorsa descritta
- tutte le proprietà, ad esclusione dell'elemento *rights*, possono essere ripetute⁹
- ogni valore è una stringa letterale, a cui può essere associata una lingua
- il collegamento tra un record DC e la risorsa descritta può essere realizzato codificando l'URI della risorsa come valore dell'elemento Identificatore DC, anche se non obbligatorio.

Il tracciato proposto per l'interoperabilità con l'aggregatore è il Dublin Core nella forma **simple**, al quale sono stati aggiunti altri elementi contenuti in uno schema definito ad-hoc nell'ambito del progetto Fare Conoscere.

⁸ La traduzione italiana di riferimento della versione 1.1 del "Dublin Core Metadata Element Set" è curata dall'[ICCU](#) (Istituto Centrale per il Catalogo Unico delle Biblioteche Italiane e per le Informazioni Bibliografiche), mentre quella francese è curata da [INRIA](#) (Institut National de Recherche en Sciences et Technologies du Numérique)

⁹ Esempio:

```
<dc:subject>cancellata</dc:subject>
<dc:subject>costruito storico</dc:subject>
<dc:subject>medio evo</dc:subject>
```

Per i dettagli sullo schema del tracciato di interoperabilità si rimanda al documento 3.2 Definizione degli strumenti di sharing, Allegato 1.

5.2 Regole di nomenclatura dei file

I partner che conferiranno i dati attraverso un file statico dovranno rispettare le seguenti regole per la nomenclatura dei file, in modo da fornire una serie di macro-elementi per la corretta alimentazione dell'aggregatore.

Il nome del file potrebbe essere strutturato nei seguenti cinque elementi, concatenati tra loro e separati dal carattere "-":

NomePartner-NomeSistema-Ente-Tipo-Data.XML

Non utilizzare caratteri con accento nei nomi dei file.

Di seguito il dettaglio degli elementi:

- **NomePartner:** denominazione codificata dell'ente partner che conferisce i dati
 - Regione Piemonte: REGPIE
 - Regione Valle d'Aosta: REGVDA
 - Regione Liguria: REGLIG
 - Ville de Nice: VDNICE
- **NomeSistema:** denominazione del sistema gestionale da cui derivano i dati
 - Mèmora: MEM
 - Biblioteca Digitale Ligure: BDL
 - Giornali del Piemonte: GDP
 - Catalogo dei Beni Culturali: CBC
 - Syracuse: SYR
- **Ente:** denominazione dell'ente afferente al NomeSistema a cui si riferiscono i dati, senza caratteri di separazione. Ad esempio, per Mèmora l'elemento è rappresentato dalla denominazione dell'ente titolare dei dati (es. AccademiaAlbertina)
- **Tipo:** fornisce l'elemento di dettaglio rispetto all'Ente, necessario per identificare la macrocategoria di dati contenuti nel file.
 - per Mèmora è rappresentato dalla tipologia di scheda a cui si riferiscono i dati:
 - OA Opere d'arte
 - S Stampe
 - D Disegni
 - F Fotografie
 - ARC Archivio storico
 - BDM Beni demontoantropologici materiali
 -
 - per Giornali del Piemonte è rappresentato dalla data di edizione della testata nel formato aaammgg (es. 19171225)
- **Data:** data di creazione del file XML, funzionale alla gestione degli aggiornamenti sul sistema aggregatore

Si forniscono di seguito due esempi di nome file che derivano dalle regole precedenti, per dati riferiti a Mèmora e per dati derivanti da Giornali del Piemonte:

- REGPIE-MEM-AccademiaAlbertina-OA-20200713.XML: dati conferiti da Regione Piemonte, derivanti dal sistema Mèmora, riferiti all'Accademia Albertina, schede OA Opere d'arte, file generato il 13/7/2020

- REGPIE-GDP-Gazzetta del Popolo-19171225-20200713.XML: dati conferiti da Regione Piemonte, derivanti dal sistema Giornali del Piemonte, riferiti alla Gazzetta del Popolo, edizione del 25/12/1917, file generato il 13/7/2020

5.3 Modalità di interscambio

L'analisi tecnica ha evidenziato che la Biblioteca Digitale Ligure e il Portale della Ville de Nice forniscono i dati tramite server OAI-PMH, mentre Mèmora e il Catalogo dei Beni Culturali della Valle d'Aosta mettono a disposizione dati statici, rispettivamente in XML e JSON.

Come modo di interscambio, si potrà utilizzare quindi uno dei seguenti metodi:

- OAI-PMH
- file statico (XML, JSON)

In entrambi i casi, i sistemi di gestione dovranno fornire file strutturati secondo lo schema del tracciato di interscambio, riportato in precedenza, fatta eccezione della Regione Liguria, che fornirà i record in un formato proprietario. La Regione Liguria dovrà comunque fornire un mapping tra il proprio formato e il tracciato di interoperabilità.

I sistemi di gestione che forniscono i dati secondo il protocollo OAI-PMH dovranno essere interrogabili tramite il protocollo HTTP/HTTPS.

I sistemi di gestione che non implementano un server OAI, potranno fornire i dati all'aggregatore secondo almeno una delle seguenti modalità:

- URL di download contenente il file statico: il partner crea il file XML su una propria area e l'aggregatore lo recupera via HTTP/FTP
- upload - il partner carica il file statico sull'aggregatore: il partner invia via HTTP/FTP il file sull'area che l'aggregatore metterà a disposizione.

Per i dettagli sulle modalità di interscambio si rimanda al documento 3.2 Definizione degli strumenti di sharing, Sezione 5,

5.4 Multilinguismo

I sistemi di gestione disponibili presso i partner mettono a disposizione risorse principalmente in una delle seguenti due lingue: francese o italiano. Nell'ambito del progetto, la traduzione dei campi a testo libero non è prevista dai partner, per cui dovrebbe essere demandata totalmente all'aggregatore. I partner dovranno comunque specificare la lingua di ciascun record, come definito nel tracciato di interscambio.

Per i dettagli sul multilinguismo si rimanda al documento 3.2 Definizione degli strumenti di sharing, Sezione 7.

5.5 Politiche di aggiornamento dei dati

La frequenza di aggiornamento concordata è trimestrale. Il dettaglio rispetto alle modalità ed il calendario di aggiornamento dei dati è descritto nel documento 3.2 Definizione degli strumenti di sharing Sezione 10.

L'aggiornamento consiste nel trasmettere l'intero insieme dei metadati relativi alle risorse (solo per quelle che hanno subito aggiornamenti), perché la trasmissione dei soli aggiornamenti richiederebbe lo sviluppo di ulteriori sistemi di controllo sulle variazioni.

6. Raccomandazioni tecniche per la valorizzazione e la promozione

Come sintesi del documento di analisi tecnica emerge quanto segue:

Relativamente alla valorizzazione, i partner italiani rispettano la normativa italiana in materia di cultura, che permette la valorizzazione delle risorse tramite accesso libero alle risorse e partecipazione a bandi per la loro valorizzazione. A livello francese, questa informazione non è disponibile.

Si suggerisce di attenersi alla normativa descritta nel documento di Analisi Tecnica. In aggiunta, si suggerisce di seguire le indicazioni definite nell'ambito del progetto PITEM PA.CE. Far Conoscere per la valorizzazione delle risorse digitali.

Bibliografia

[1] Aquilanti, Francesco, Francesco Baldi, and Massimo Pistacchi. "Verba Manent Teoria e prassi della digitalizzazione dei documenti sonori e video della Discoteca di Stato-Museo dell'audiovisivo (prima parte)." *Digitalia 2* (2006): 131-148.

Rapport de Recommandations Techniques

Région Piémont

Délégués

Istituto di Informatica e Telematica - CNR

CSI Piemonte

Glossaire

Ressource d'origine

Ressource culturelle telle que décrite dans le système natif, comme par exemple une peinture.

Copie numérique

Résultat d'un processus de numérisation d'une ressource d'origine. Plusieurs copies numériques peuvent correspondre à une ressource d'origine, par exemple différentes photos prises de la même peinture.

Métadonnée

Attribut qui spécifie une propriété de la ressource d'origine ou de la copie numérique (par exemple le titre d'une ressource d'origine ou le format de la copie numérique).

Record

L'ensemble de toutes les métadonnées d'une ressource d'origine et des copies numériques associées.

Ensemble de métadonnées pour l'interopérabilité

Schéma des collections de record fournies à l'agrégateur.

Format de l'ensemble de métadonnées pour l'interopérabilité

Un des formats suivants : xml, json, csv, ...

Sommaire

[1. Introduction](#)

[2. Typologie de ressources à numériser](#)

[3. Recommandations techniques pour la numérisation](#)

[3.1 Images](#)

[3.2 Audio et Vidéo](#)

[4. Recommandations techniques pour le catalogage](#)

[5. Interopérabilité](#)

[5.1 Ensemble des métadonnées pour l'échange](#)

[5.2 Règles de nomenclature des fichiers](#)

[5.3 Modalité d'échange](#)

[5.4 Multilinguisme](#)

[5.5 Politiques de mise à jour des données](#)

[6. Recommandations techniques pour la valorisation et la promotion](#)

[Bibliographie](#)

1. Introduction

Ce document décrit les recommandations techniques que chaque partenaire devrait considérer pour la numérisation, le catalogage et la valorisation de ses propres ressources culturelles et l'interopérabilité de ses systèmes de gestion. Les recommandations proposées sont à considérer indicatives, puisqu'elles dépendent des ressources à numériser, à évaluer cas par cas selon le projet considéré.

Du document d'analyse technique il ressort que les systèmes de gestion mis à disposition par les partenaires dans leur rôles de fournisseurs de données pour l'agrégateur sont :

- Biblioteca Digitale Ligure (BDL) (Regione Liguria)
- Giornali del Piemonte (Regione Piemonte)
- Mèmora (Regione Piemonte)
- Catalogo dei Beni Culturali (Regione Valle d'Aosta)
- Syracuse (Ville de Nice).

Les ressources qui seront mises à disposition pour l'utilisation à travers de l'agrégateur devront respecter les exigences suivantes :

1. au moins une copie numérique doit être associée à chaque ressource ;
2. il est possible d'accéder à la copie numérique à travers d'un URI ;
3. la ressource est connectée au moins à un des thèmes du projet PITEM Pa.CE ;
4. la ressource numérique est utilisable avec une licence qui en permet l'utilisation ;
5. la responsabilité de la ressource est en charge à l'institution qui est titulaire de la donnée, c'est-à-dire qui détient le bien.

2. Typologie de ressources à numériser

Pour analyser les exigences pour les nouvelles numérisations on a demandé aux partenaires d'indiquer dans le questionnaire non seulement le nombre des nouvelles ressources qu'ils prévoient de numériser, mais aussi leur typologie.

Des données reçues d'une façon préliminaire, il ressort que les typologies identifiées sont : photos, cartes, dessins, plaques, diapos, audio, vidéo, journaux, livres, périodiques. On peut donc conclure que les formats pour lesquels il est nécessaire d'élaborer des lignes directrices pour la numérisation sont en essentiel quatre : images, audio, vidéo et texte.

Par *image* on entend généralement un artefact bidimensionnel qui représente un sujet. Dans le cas d'une photo d'un objet d'art (par exemple une statue), la photo est la représentation bidimensionnelle d'un objet tridimensionnel. Dans ce rapport on entend soit images produites à travers instruments optiques (appareils photo), c'est à dire photos ou diapos, soit toute représentation bidimensionnelle telle que carte, graphique, dessin, peinture, panneau.

Les images en mouvement (avec ou sans audio) sont comprises dans le format *vidéo*, tandis que pour *audio* on entend enregistrement de sons sans images.

Par *texte* on considère un objet qui peut être lu, donc pour nos bouts on comprend dans le format textuel documents, livres, journaux, périodiques.

3. Recommandations techniques pour la numérisation

Dans le cadre du patrimoine culturel on trouve différentes typologies de ressources culturelles, chacune caractérisée par son propre format, par le matériel et par le type de document. Pour numériser ces ressources il est nécessaire présumer que le document soit plat et que la présence d'éventuelles rugosités ou plis ne compromette pas le processus de numérisation.

Indépendamment de la typologie des ressources considérées, généralement il existe au moins les trois types suivants de mémorisation sur disc de la même ressource :

- **master** (pour la conservation) - la représentation doit être effectuée pour garantir une qualité très haute de la ressource ; une ressource master doit maintenir le plus grand nombre d'informations possible. Ce type de mémorisation n'est pas partagé avec l'agrégateur, donc chaque partenaire peut utiliser ses propres lignes directrices pour la mémorisation master. Toutefois, dans un souci d'exhaustivité, on propose des recommandations techniques aussi pour ce type de mémorisation ;
- **dissémination** (destinée à l'utilisation sur navigateur web) - la représentation doit assurer le téléchargement rapide de la part du navigateur, mais il ne doit pas être trop de mauvaise qualité. Donc, dans ce cas, on suggère une numérisation qui assure une qualité moyenne de la ressource.
- **ongle du pouce/échantillon** - représentation à basse résolution, qui permet d'avoir une prévisualisation de la ressource qui peut être inclus, par exemple, dans une page qui contient le résultat d'une recherche effectuée par l'utilisateur.

Du document *Inventaire Sites Web* il ressort que les modèles de diffusion les plus utilisés par les Digital Libraries for Cultural Heritage (DLCH) présentent les résultats en utilisant seulement les niveaux ongle du pouce et dissémination, et non le niveau master. En général le niveau ongle du pouce est produit automatiquement par la DLCH de l'image de dissémination. Donc, pour l'interopérabilité avec l'agrégateur, en théorie il suffirait définir seulement les lignes directrices de numérisation pour l'image de dissémination, en laissant les partenaires libres de numériser les images mater selon leurs exigences. Toutefois, dans un souci d'exhaustivité, le présent document offre des lignes directrices pour la numérisation des images master et ongle du pouce.

Chaque typologie de numérisation est caractérisée par des exigences minimales de numérisation. On suggère aux partenaires de respecter les lignes directrices indiquées dans ce document, pour assurer une homogénéité dans la présentation des résultats sur l'agrégateur.

3.1 Images

Le processus de numérisation d'une image est une transformation de l'image continue (analogique) dans une matrice de nombres entiers. Le processus s'effectue dans deux phases successives : l'échantillonnage et la quantification. L'échantillonnage consiste en diviser l'image dans un réticule de points (échantillons), les pixels. À parité de taille de l'image, le nombre de pixels fournit indications sur la résolution de l'image, c'est à dire sur la fréquence d'échantillonnage spatial appliquée. La quantification approxime le valeur des échantillons à un nombre déterminé de

valeurs. En général la quantification consiste en utiliser une palette de couleurs pour la représentation des valeurs des échantillons :

- les images noir et blanc sont quantifiées avec 1 bit/pixel
- les images en gris sont quantifiées typiquement sur 256 niveaux (1 byte/pixel)
- les images en couleur se décomposent sur plusieurs canaux (RGB, CMYB, etc.) et chaque canal est quantifié séparément (atteignant 3/4 byte pour pixel).

Dans le processus de numérisation des images, les erreurs plus typiques à éviter sont :

- image verticale numérisée en mode paysage et non portrait (ou vice versa)
- rotation de l'image d'origine
- image à l'envers
- problèmes de lumière avec le scanner
- image coupé
- poussière sur la plaque du scanner¹⁰.

De la synthèse du document d'analyse technique il ressort que :

Pour ce qui concerne la numérisation, tous les partenaires effectueront les numérisations à une résolution minimale des images master (300 ppp) pour l'échelle de couleur : 24 bit pour les images en couleur et 8 bit pour les images en noir et blanc. Pour les images de dissémination (consultation sur le web), il n'est pas possible de définir une valeur minimale de résolution, car il y a des typologies d'images différentes, chacune avec une résolution différente. Pour ce qui concerne la recherche texte, Giornali del Piemonte utilise le format PDF avec ROC intégré.

Master

L'objectif de l'image master est de maintenir toutes les informations, sans en perdre aucune. Pour identifier la correcte résolution de l'image, on suggère d'utiliser techniques d'interpolation, comme spécifié ici en suite. L'interpolation est le processus que, en partant des pixels réels présents dans l'image, estime les pixels non connus. Il y a plusieurs techniques d'interpolation¹¹, que différents programmes de numérisation utilisent dans l'élaboration des images numérisées. Pour trouver la résolution optimale pour la numérisation d'une image, dans un premier temps, on essaye d'effectuer deux numérisations de la même image, respectivement à une résolution $R1 = X$ et $R2 = X + \square$. Ensuite, on applique à l'image avec la résolution $R1$ une technique d'interpolation pour obtenir la résolution $R2^*$. On compare ensuite les images avec résolutions $R2$ et $R2^*$: s'il n'y a pas de différences entre les deux images, alors la résolution optimale pour la numérisation est $R1$; si on note des différences, on continue avec une ultérieure numérisation de l'image avec une résolution $R3 = R2 + \square$, on effectue l'interpolation de $R2$ pour obtenir $R3^*$ et ainsi de suite jusqu'à obtenir la résolution optimale.

On suggère de numériser les images en haute résolution dans le format TIFF compressé et non compressé, avec profondeur de numérisation en couleur (24 bit) ou en tons de gris (8 bit) avec une résolution finale qui dépend de la typologie du matériel. En général, la résolution finale est environ de 300 ppp.

¹⁰ Fonte:

<https://www.sabap-abruzzo.beniculturali.it/wp-content/uploads/Normativa-per-lacquisizione-digitale-delle-immagini-fotografiche.pdf>

¹¹ <http://www.di.univr.it/documenti/OccorrenzaIns/matdid/matdid358544.pdf>

Dissémination

Dans ce cas l'objectif est l'utilisation sur le Web. L'image de dissémination peut être générée avec un sous-échantillonnage de l'image master, à travers de logiciels dédiés de manipulation d'images, pour obtenir une résolution plus basse, utilisable pour l'utilisation sur le Web.

En général, la résolution optimale pour les images de dissémination dépend du type de dispositif utilisé par l'utilisateur final. Les Figures 1, 2 et 3 montrent les résolutions de l'écran les plus diffusées en 2020, pour typologie de dispositif (de bureau, tablette et smartphone)¹². On suggère d'utiliser, pour l'image de dissémination, la résolution la plus diffusée pour ordinateur de bureau (1366 px x 768 px). Les autres résolutions peuvent être produites avec sous-échantillonnage de la résolution proposée.

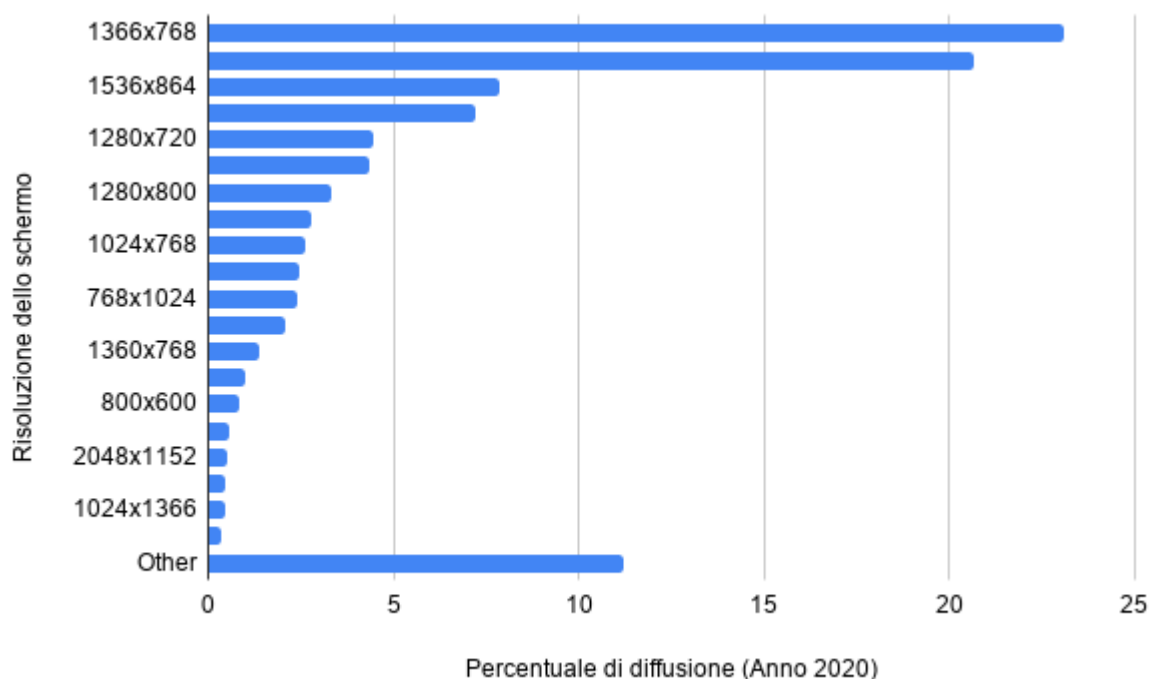


Fig. 1 Pourcentage de diffusion des différentes résolutions d'écran pour le 2020 pour ordinateur du bureau.

¹² Fonte *Stat Counter Global Stats*: <https://gs.statcounter.com/screen-resolution-stats/>

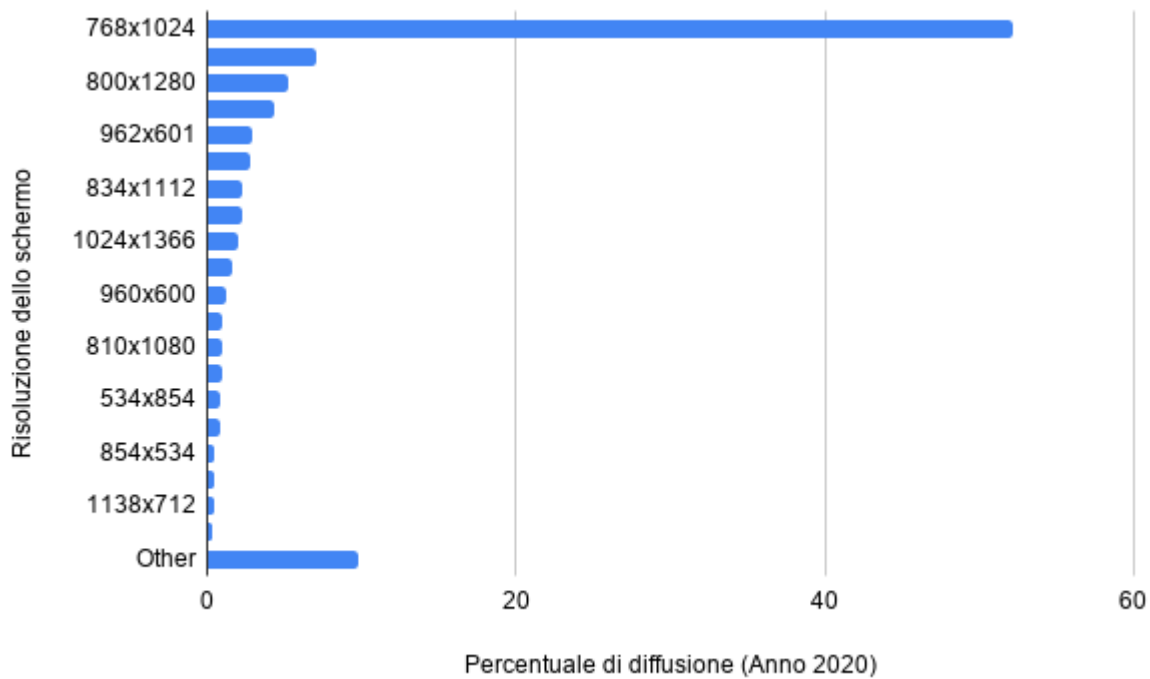


Fig. 2 Pourcentage de diffusion des différentes résolutions d'écran pour le 2020 pour tablette.

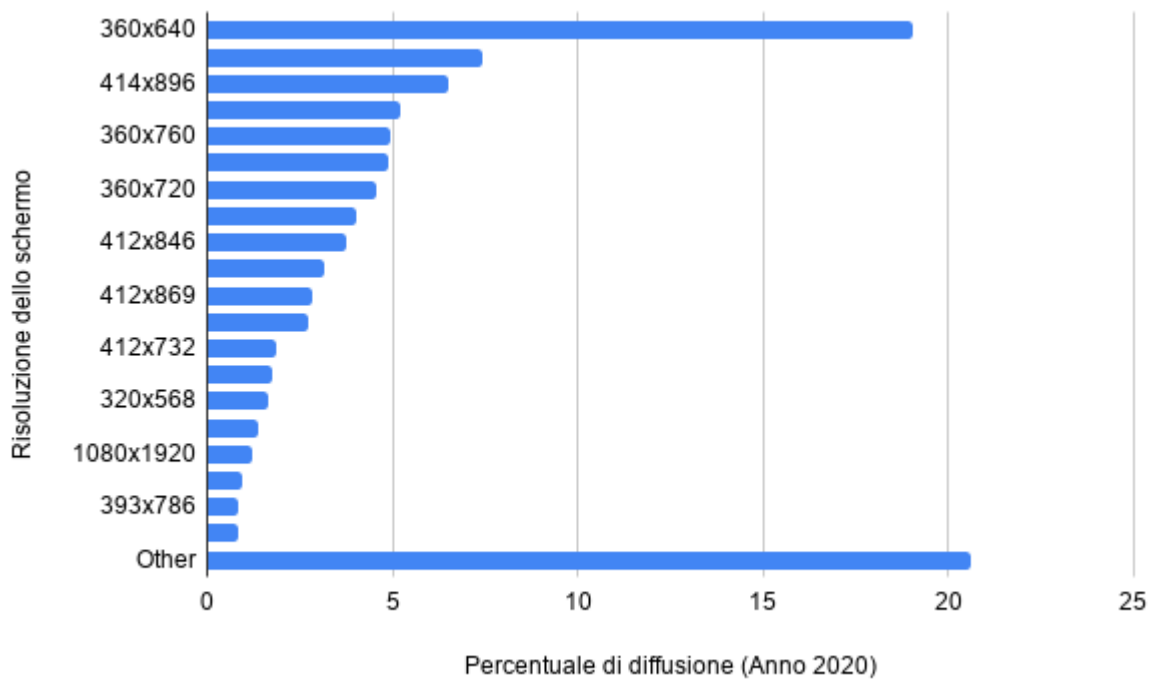


Fig. 3 Pourcentage de diffusion des différentes résolutions d'écran pour le 2020 pour smartphone.

Ongle de pouce

Numérisation à basse résolution qui contient une prévisualisation de l'image ou d'un détail qui permet à l'œil humain d'en comprendre le contenu. L'agrégateur extraira l'image ongle de pouce

de l'image de dissémination. Si cette image n'est pas disponible, l'agrégateur procédera à insérer une image ongle de pouce de défaut ou une image de substitution.

3.2 Audio et Vidéo

Même pour les ressources audiovisuelles, en général on crée une copie master pour la conservation et des copies compressées pour la consultation.

Le ICBSA¹³ (Istituto Centrale per i Beni Sonori e Audiovisivi - Institut Central pour les Biens Sonores et Audiovisuels) a publié ses propres lignes directrices pour la numérisation de documents sonores et vidéo [1]. La transcription en numérique est effectuée en adoptant les indications de la International Association of Sound and Audiovisual Archives (IASA)¹⁴, conformément au standard BWF produit par la European Broadcasting Union (EBU)¹⁵ en collaboration avec la Audio Engineering Society (AES)¹⁶, en acceptant les indications sur la résolution de l'Unesco sur la conservation du numérique et les indications des activités du projet Presto (Preservation Technology for European Archives), sur la préservation des archives analogiques audio, vidéo et cinéma, guidé par BBC, RAI et Institut National de l'Audiovisuel avec ACS et ITC/IRST en qualité de partenaires techniques pour la partie audio.

Il est recommandé, pour les copies master à haute qualité, d'utiliser le standard européen Broadcast Wave Format (BWF), une extension du format WAV qui permet d'ajouter dans l'en-tête du fichier quelques métadonnées, sans la nécessité d'un lecteur audio spécifique pour la reproduction.

L'échantillonnage du matériel analogique est à 48 kHz 24 bit stéréo dans les cas normales, et à 96 kHz 24 bit dans les cas où le matériel doit être restauré. Le matériel nativement numérique (par exemple CD) est numérisé avec l'échantillonnage d'origine.

Les copies compressées (pour réduire l'espace de mémorisation et le débit de transmission nécessaires) servent pour la consultation du matériel à l'intérieur de l'Institut ou par Internet. Le débit binaire et l'algorithme de codification peuvent varier sur la base des progrès technologiques, mais des valeurs utilisables sont 256 kbit/s avec codification MPEG 1 Layer 3 pour la copie à haute qualité, et 32 kbit/s avec codification MPEG 1 Layer 3 pour celles de qualité plus basse.

4. Recommandations techniques pour le catalogage

De la synthèse du document d'analyse technique il ressort que :

Pour ce qui concerne le catalogage, les différents partenaires utilisent standards et lignes directrices différentes : pour les fiches de catalogage Biblioteca Digitale Ligure et SBNWeb utilisent ICCU, tandis que Mèmora et Catalogo dei Beni Culturali della Valle d'Aosta utilisent ICCD. La Ville de Nice utilise UNIMARC. Il n'y a pas d'unanimité non plus sur les standards de catalogage utilisés : seulement Mèmora et la Biblioteca Digitale Ligure utilisent le Dublin Core.

¹³ <http://www.icbsa.it/>

¹⁴ <https://www.iasa-web.org/>

¹⁵ <https://www.ebu.ch/home>

¹⁶ <http://www.aes.org/>

Pour l'interopérabilité entre les différents systèmes de gestion, il n'est pas nécessaire que les ressources numériques soient cataloguées d'une façon uniforme et que tous les systèmes utilisent le même standard de catalogage ou les mêmes lignes directrices. La chose la plus importante est que tous les systèmes utilisent le même ensemble des métadonnées (décrit dans la section suivante). Donc, pour le catalogage des ressources, on suggère d'utiliser un des standards décrits dans le document d'Analyse Technique.

5. Interopérabilité

5.1 Ensemble des métadonnées pour l'échange

Dans ce chapitre on décrit l'ensemble des métadonnées pour l'interopérabilité qu'on propose d'utiliser pour partager les ressources que les systèmes des partenaires vont à identifier pour partager et utiliser à travers de l'agrégateur.

On propose donc un standard qui soit au même temps flexible, capable de représenter objets différents, reconnu comme standard international et qui permet l'exportation des données d'une façon simple.

Le standard qui répond à toutes ces exigences est le Dublin Core¹⁷, structuré dans deux formes : **simple** et **qualified**. La première forme se compose de 15 éléments de base, la deuxième intègre les éléments de base avec des nouveaux balises pour offrir une description plus détaillée du produit.

Les caractéristiques principales du standard identifié sont :

- la **facilité** d'utilisation, car le langage est compréhensible soit par spécialistes, soit par novices ;
- l'extensibilité qui permet d'ajouter aux quinze éléments de base des nouveaux éléments ;
- l'utilisation d'une **sémantique claire et universellement acceptée**, car il a été créé par des experts de différents nations et il est orienté à assurer l'**interopérabilité** avec d'autres profils de métadonnées.

La logique du standard Dublin Core peut être ainsi synthétisée :

- un record DC se compose d'une ou plusieurs propriétés, représentée par l'élément et les valeurs associés
- chaque propriété est un attribut de la ressource décrite
- toutes les propriétés, avec l'exclusion de l'élément *rights*, peuvent être répétées¹⁸
- chaque valeur est une chaîne de lettres, à laquelle on peut associer un langage
- la connexion entre un record DC et la ressource décrite peut aussi être réalisée codant l'URI de la ressource tel que valeur de l'élément identificateur DC, même s'il n'est pas obligatoire.

¹⁷ La traduction de référence de la version 1.1 du "Dublin Core Metadata Element Set" est supervisée par [ICCU](#) (Istituto Centrale per il Catalogo Unico delle Biblioteche Italiane e per le Informazioni Bibliografiche), tandis que la française est supervisée par [INRIA](#) (Institut National de Recherche en Sciences et Technologies du Numérique)

¹⁸ Exemple:

```
<dc:subject>cancellata</dc:subject>
<dc:subject>costruito storico</dc:subject>
<dc:subject>medio evo</dc:subject>
```

L'ensemble de métadonnées proposé pour l'interopérabilité avec l'agrégateur est le Dublin Core dans la forme **simple**, auquel on a ajouté des éléments compris dans un schéma défini ad-hoc dans le cadre du projet Faire Connaître.

Pour détails sur le schéma de l'ensemble de métadonnées pour l'interopérabilité, veuillez consulter le document 3.2 Définition des outils de partage, Annexe 1.

5.2 Règles de nomenclature des fichiers

Les partenaires qui enverront leurs données à travers d'un fichier statique devront respecter les règles suivantes pour nommer les fichiers, pour fournir un ensemble de macro-éléments pour la correcte alimentation de l'agrégateur.

Le nom du fichier pourrait être structuré dans les cinq éléments suivants, connectés entre eux et séparés par le caractère "-":

NomPartenaire-NomSysteme-Institution-Type-Date.XML

N'utilisez pas des caractères avec accent dans les noms des fichiers.

La description de détail des éléments est la suivante:

- NomPartenaire : dénomination codifiée de l'institution partenaire qui envoie les données
 - Regione Piemonte: REGPIE
 - Regione Valle d'Aosta: REGVDA
 - Regione Liguria: REGLIG
 - Ville de Nice: VDNICE
- NomSysteme : dénomination du système de gestion duquel résultent les données
 - Mèmora: MEMORA
 - BDL:
 -
- Institution: dénomination de l'institution associée au NomSysteme auquel se réfèrent les données, sans caractères de séparation, par exemple pour Mèmora il est représenté par la dénomination de l'institution titulaire des données (par ex. AccademiaAlbertina)
- Type : il fournit un élément de détail par rapport à l'institution, nécessaire pour identifier la macro catégorie de données contenues dans le fichier.
 - pour Mèmora il est représenté par la typologie de fiche à laquelle se réfèrent les données :
 - OA Oeuvres d'art
 - S Estampes
 - D Dessins
 - F Photos
 - ARC Archive historique
 - BDM Biens matériaux démographiques ou anthropologiques
 -
 - pour Giornali del Piemonte il est représenté par la date d'impression de l'organe de presse dans le format aaaammjj (es. 19171225)
- Date : date de création du fichier XML, fonctionnel à la gestion de la mise à jour sur le système agrégateur

Ici en suite on présente deux exemples de nom de fichiers qui dérivent des règles précédentes, pour données référées à Mèmora et pour données qui dérivent de Giornali del Piemonte:

- REGPIE-MEM-AccademiaAlbertina-OA-20200713.XML : données envoyés par Regione Piemonte, provenant du système Mèmora, référés à l'Accademia Albertina, Fiches OA Oeuvres d'art, fichier généré le 13/7/2020
- REGPIE-GDP-GazzettadelPopolo-19171225-20200713.XML : données envoyées par Regione Piemonte, provenant du système Giornali del Piemonte, référés à la Gazzetta del Popolo, édition du 25/12/1917, fichier généré le 13/7/2020

5.3 Modalité d'échange

L'analyse technique a souligné que la Biblioteca Digitale Ligure et le Portail de la Ville de Nice envoient données à travers serveur OAI-PMH, alors que Mèmora et le Catalogo dei Beni Culturali de la Vallée d'Aoste mettent à disposition données statiques, respectivement en XML et JSON.

Comme modalité d'échange on pourra donc utiliser un des méthodes suivantes :

- OAI-PMH
- fichier statique (XML, JSON)

Dans les deux cas, les systèmes de gestion doivent envoyer fichiers structurés selon le schéma pour l'échange de l'ensemble de métadonnées déjà présenté, à l'exception de la Regione Liguria, qui enverra les records dans un format propriétaire. La Regione Liguria devra de toute façon présenter la correspondance entre son propre format et l'ensemble de métadonnées.

Les systèmes de gestion qui envoient les données selon le protocole OAI-PMH devront être consultables à travers le protocole HTTP/HTTPS

Les systèmes de gestion qui n'utilisent un serveur OAI pourront envoyer les données à l'agrégateur selon au moins une des modalités suivantes :

- URL de téléchargement avec le fichier statique : le partenaire crée le fichier XML dans son système et l'agrégateur le récupère par HTTP/FTP
- téléchargement - le partenaire charge le fichier statique sur l'agrégateur : le partenaire envoie par HTTP/FTP le fichier sur le dossier que l'agrégateur mettra à disposition.

Pour d'autres détails sur les modalités d'échange, veuillez consulter le document 3.2 Définition des outils de partage, Section 5.

5.4 Multilinguisme

Les systèmes de gestion disponibles chez les partenaires visualisent les ressources principalement dans une des deux langues : français ou italien. Dans le cadre du projet, la traduction des champs avec texte libre n'est pas prévue par les partenaires, donc elle devrait être confiée totalement à l'agrégateur. Les partenaires devront de toute façon spécifier la langue de chaque record, tel que défini dans l'ensemble de métadonnées pour l'échange.

Pour les détails sur le multilinguisme, veuillez consulter le document 3.2 Définition des outils de partage, Section 7.

5.5 Politiques de mise à jour des données

La fréquence de mise à jour concertée est tous les trois mois. Le détail par rapport aux modalités et au calendrier de mise à jour des données est décrit dans le document 3.2 Définition des outils de partage, Section 10.

La mise à jour consiste en envoyer l'ensemble entier des métadonnées des ressources (seulement pour celles qui ont été mises à jour), parce que la transmission des seules données modifiées impliquerait le développement d'autres systèmes de contrôle sur les variations.

6. Recommandations techniques pour la valorisation et la promotion

De la synthèse du document d'analyse technique il ressort que :

Relativement à la valorisation, les partenaires italiens respectent la législation italienne sur la culture, qui permet la valorisation des ressources avec accès libre aux ressources et participation à appels d'offre pour les contributions pour les valoriser. Au niveau français cette information n'est pas disponible.

On suggère de respecter la législation écrite dans le document d'Analyse Technique. En plus, on suggère de suivre les indications définies dans le cadre du PITEM Pa.C.E. Faire connaître pour la valorisation des ressources numériques.

Bibliographie

[1] Aquilanti, Francesco, Francesco Baldi, and Massimo Pistacchi. "Verba Manent Teoria e prassi della digitalizzazione dei documenti sonori e video della Discoteca di Stato-Museo dell'audiovisivo (prima parte)." *Digitalia 2* (2006): 131-148.