REGIONE MARCHE

Agenda Digitale Marche

DIGiP

DIGital Preservation

Polo di conservazione regionale

Polo di conservazione regionale Marche DigiP

P.F. Informatica e Crescita Digitale
Dirigente Dott.ssa Serenella Carota

Il Polo di conservazione Marche DigiP: modello e prospettive future

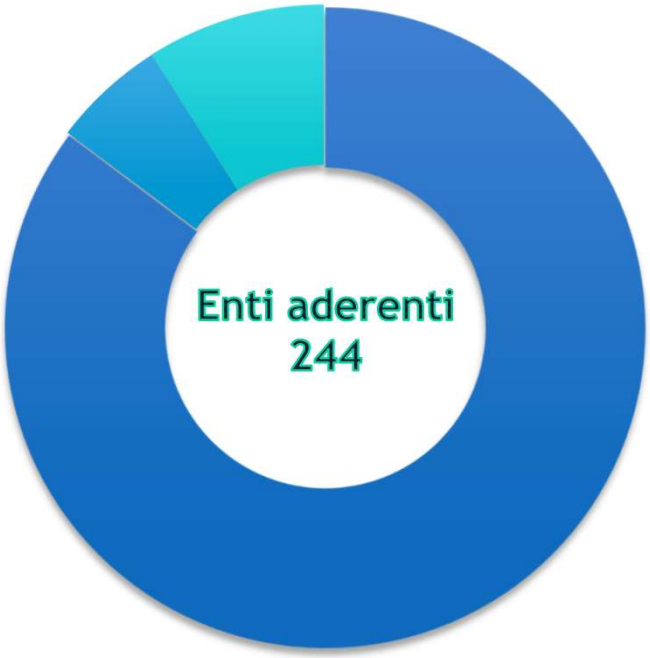
- I numeri del Polo Marche DigiP
- Tipo di oggetti digitali conservati
- Strutturazione documentale
- Aree funzionali
- Architettura (livello di presentazione, business logic, persistenza)
- Modalità di invio dei pacchetti in conservazione
- Controlli di qualità (Quality Assurance)
- Riconoscimento e validazione dei formati
- Protocollazione
- Caratteristiche dei documenti conservati e personalizzazione
- Metadati
- Prospettive future e possibili miglioramenti

Il Polo Marche DigiP

- Il Polo di conservazione Marche DigiP è stato approvato con legge regionale n. 3/2015 e messo a disposizione di tutte le Pubbliche Amministrazioni del territorio regionale con specifica convenzione. Al fine di incentivare e implementare un processo conservativo significativo e di qualità, la Regione Marche si avvale di un Comitato scientifico specialistico per il Polo di Conservazione, ha definito un tavolo regionale utilizzatori per acquisire feedback sulle attività del Polo, fornisce servizi di assistenza ed help desk per gli enti produttori e mette a disposizione anche software di protocollo e gestione documentale e di gestione degli atti digitali già interoperabili con il Polo. Sono inoltre previsti processi di verifica della qualità di quanto versato in base ad accordi specifici con i singoli enti

I numeri del Polo Marche DigiP

Al 31/12/2020 i documenti digitali conservati da Marche DigiP sono più di **23 milioni** (23.096.846)



Le Pubbliche Amministrazioni del territorio regionale attivano il servizio di conservazione con specifica convenzione. Dei **244** enti aderenti che hanno avviato l'iter di adesione al servizio:

- **202** hanno sottoscritto la convenzione
- **187** hanno avviato il servizio di conservazione

Lo stato delle adesioni, aggiornato ogni 15 giorni, è pubblicato sul sito del Polo al seguente link https://www.regione.marche.it/Regione-Utile/Agenda-Digitale/Polo-di-conservazione-regionale

Tipo di oggetti digitali conservati

- Documento informatico, documento amministrativo informatico, (ad esempio decreti, determine, delibere, documenti protocollati, fatture, documenti interni all'ente, referti del fascicolo sanitario elettronico). In particolare il sistema è configurabile per accogliere ogni tipologia di documento informatico, fascicolo informatico, serie documentale

Strutturazione documentale

- Unità Documentaria: aggregato logico costituito da un Documento principale e d eventuali Allegati/Annessi/Annotazioni. Se l'Unità Documentaria è costituita dal solo Documento principale questo può essere composto sia da file che da soli metadati. L'Unità Documentaria è accompagnata da un file XML o indice di versamento, il quale riporta tutti i metadati che attestano i vari contesti di produzione originaria o riuso attraversati dal documento digitale nel corso della sua esistenza. I pacchetti informativi così formati, trasmessi da un ente produttore al sistema di conservazione DigiP sono denominati pacchetti informativi di versamento (SIP) i quali, una volta validati, sono trasformati in pacchetti informativi di archiviazione (AIP)
- Unità Archivistica: individua l'unità minima indivisibile di un fondo archivistico che può aggregare più documenti fisicamente contigui o essere costituita da una singola unità documentaria

Strutturazione documentale

- Aggregazione Documentale Informatica: insieme di documenti informatici o insieme di fascicoli informatici riuniti per caratteristiche omogenee, in relazione alla natura e alla forma dei documenti o in relazione all'oggetto e alla materia o in relazione alle funzioni dell'ente
- Il versamento di un'Unità Archivistica o di un'Aggregazione Documentale Informatica viene considerato come versamento di un SIP finalizzato alla creazione dei pacchetti informativi destinati alla conservazione (AIC, Archival Information Collection). In termini operativi tale evidenza documentale è costituita da un file indice xml (SIP di soli metadati) che funge da file indice dei documenti contenuti nell'Unità Archivistica. In particolare il file indice xml è costituito da sezioni in cui sono riportate informazioni fondamentali relative all'Unità Archivistica e una sotto-sezione <IndiceDocumenti> che contiene l'elenco degli identificativi delle Unità Documentarie (AIP) appartenenti a quella particolare Unità Archivistica

DigiP è composto da 6 aree funzionali:

- **INGEST** è l'area funzionale costituita dall'insieme dei processi che sovraintendono l'accettazione delle risorse digitali inviate dai soggetti produttori e della loro preparazione per l'inclusione nel sistema di archiviazione
- ARCHIVAL STORAGE è la funzione gestisce l'immagazzinamento a lungo termine delle risorse digitali affidate al sistema
- DATA MANAGEMENT è l'unità di gestione dei metadati e del catalogo di ricerca
- ADMINISTRATION è l'area funzionale che raggruppa l'insieme delle funzioni rivolte alla gestione delle configurazioni del sistema, al monitoraggio, all'interazione con gli utenti, agli accordi di servizio con i produttori ed al mantenimento degli standard di archiviazione definiti
- PRESERVATION PLANNING è l'area funzionale che si occupa della progettazione della strategia di conservazione del sistema e delle sue modifiche a fronte dei cambiamenti tecnologici riguardanti gli oggetti archiviati e del mutamento dei bisogni espressi dalla Comunità di riferimento
- **ACCESS** è l'area funzionale dove si gestisce il flusso di richieste di documenti in uscita e la ricerca da parte del Consumatore

Architettura del sistema di conservazione

- Marche DigiP è un software di proprietà della Regione Marche e come tale rilasciabile con licenza Open Source
- L'architettura di erogazione del servizio è un sistema virtualizzato caratterizzato da alto grado di parallelismo e scalabilità
- Nel dettaglio, il sistema di conservazione Marche DigiP è una web application verticale la quale architettura è costituita da tre livelli principali:
  - livello di presentazione
  - livello di business logic
  - livello di persistenza

Livello di presentazione costituito da:

- una serie di interfacce web user-oriented utili alla configurazione del sistema, al controllo del processo di versamento e alla ricerca e recupero dei documenti conservati
- una serie di interfacce standard REST e SOAP per le comunicazioni con sistemi esterni (B2B)

Livello di business logic costituito da:

- servizi generali e specializzati invocati direttamente dai servizi REST per l'implementazione delle funzionalità OAIS (Administration, Preservation Planning, Ingest, Access)
- un processo di Ingest che prevede l'impiego di un sistema a code realizzato con un message broker ad alte prestazioni (RabbitMQ). Il design a code garantisce un'elevata robustezza e disponibilità anche grazie a funzionalità di ripristino dell'elaborazione in caso di errori
- un servizio di PreIngest realizzato attraverso un modulo dedicato in Spring Boot e che si interfaccia al resto del sistema attraverso RabbitMQ, pensato per la pre-elaborazione di pacchetti prima di sottoporli al processo di Ingest. Ad oggi impiegato per l'importazione dei pacchetti di distribuzione provenienti da altro conservatore
- un servizio automatico di conservazione a norma del log applicativo in cui vengono raccolti e conservati per ogni giorno gli eventi e le azioni eseguite sul sistema (come ad esempio l'autenticazione degli utenti o le operazioni CRUD sulle varie entità logiche durante il processo di Ingest)

Livello di persistenza costituito da:

- storage temporanei ad alta disponibilità per l'elaborazione dei SIP versati durante la fase di Ingest
- storage di grande capacità dove vengono mantenuti inalterati i diversi IP (SIP, AIP e DIP)
- database per le ricerche dei diversi IP dove vengono mantenuti i metadati secondo l'organizzazione proposta dal modello OAIS

Funzionalità Pre-Ingest per Riversamento da altro Conservatore

- La funzionalità pre-ingest consente l'importazione dei pacchetti informativi provenienti da altro conservatore e conformi allo standard SInCRO nel sistema di conservazione DigiP, attraverso la generazione di un nuovo SIP conforme alle specifiche di versamento di DigiP e allo standard SInCRO
- La sezione è altamente configurabile a seconda dell'ente tramite XSLT. L'uso degli XSLT rende le trasformazioni customizzabili e configurabili senza bisogno di modificare il codice dell'applicativo

Modalità di invio dei pacchetti in conservazione

Il trasferimento di un pacchetto di versamento (SIP) è assicurato attraverso un canale di comunicazione sicuro che implementa gli algoritmi crittografici e assicura la verifica dell'integrità delle sequenze binarie trasmesse

DigiP permette l'invio dei pacchetti in conservazione attraverso tre diverse modalità:

- versamento manuale via Form Web
- versamento via API REST
- versamento via SFTP che permette il caricamento diretto dei pacchetti in formato .zip all'interno di una directory condivisa in rete

Versamento via Form Web e API REST

Per velocizzare i versamenti di tipo Manuale e REST, la fase di Ingest è stata suddivisa in due sotto-fasi distinte:

- una fase sincrona: la Presa in Carico
- una fase asincrona: l'Elaborazione

La Presa in Carico è una fase **sincrona** molto snella, in cui il pacchetto versato viene sottoposto esclusivamente ad una serie di controlli preliminari sulla conformità dell'indice e degli allegati (validazione XSD dell'indice, controllo dell'HASH degli allegati rispetto a quanto dichiarato nell'indice, etc.). Al termine della fase sincrona la chiamata REST restituisce una **Ricevuta di Carico** che attesta che il pacchetto è stato effettivamente preso in carico dal Sistema

L'Elaborazione è una fase **asincrona** molto più corposa in cui il pacchetto viene sottoposto ai controlli di qualità e trasformato in standard SInCRO, in base alla configurazione prevista per l'ente. Al termine di questa fase il pacchetto viene considerato archiviato o non validato (nel caso non siano superati i controlli sulla qualità). Tale fase produce una **Rapporto di Versamento**

Versamento via SFTP

- Il versamento via SFTP è completamente asincrono e permette ad un ente di versare i pacchetti all'interno di una directory personale senza doversi sincronizzare in alcun modo con DigiP. Al termine dell'elaborazione dei pacchetti i Rapporti di Versamento saranno restituiti da DigiP nella medesima directory.

Controlli di qualità (Quality Assurance)

Il processo di Quality Assurance prevede una serie di controlli quali:

- controllo di validità dell'Indice di versamento con il file schema XSD
- analisi dei formati dei file in essi contenuti
- verifica delle eventuali firme digitali con conseguente recupero dei dati dei firmatari
- controllo dell'hash dichiarato nell'indice di versamento per ogni file contenuto nel SIP
- controllo della presenza di virus per ogni file contenuto nel SIP

Riconoscimento e validazione dei formati

- L’analisi dei formati dei file in essi contenuti viene eseguita attraverso il tool FITS (File Information Tool Set) quale strumento in grado di identificare e validare formati di file, estrarre metadati incorporati all'interno di file e generare metadati tecnici in schemi XML. Funziona come un *wrapper* invocando e gestendo l'output da molti altri strumenti open source. FITS è sviluppato e mantenuto da un gruppo di ricerca dell’Università di Harvard https://projects.iq.harvard.edu/fits/home

Protocollazione del Rapporto di Versamento

- Il Rapporto di versamento è protocollato dal sistema di Protocollo dell'Ente Polo (con D.G.R. n. 56 del 23/01/2012 è stata istituita un'ulteriore Area Organizzativa Omogenea per il Polo di conservazione Marche DigiP, operante presso la P.F. Informatica e Crescita Digitale)
- La segnatura di protocollo così ottenuta rappresenta un valido riferimento temporale opponibile a terzi in quanto il sistema di protocollo che lo ha prodotto è il protocollo informatico di un ente pubblico. La marcatura temporale ottenuta per tramite del protocollo viene mantenuta in associazione con il SIP e inclusa tra i metadati del Pacchetto di archiviazione prima della necessaria apposizione della firma digitale del Conservatore. Il Rapporto di Versamento viene inoltre aggiunto al contenuto informativo del SIP originale. Questa parte del processo di acquisizione garantisce la qualità del trasferimento nei confronti di Terzi

Caratteristiche statiche/dinamiche dei documenti conservati

- Il sistema di conservazione Marche DigiP assicura il mantenimento dell'autenticità dei documenti digitali nel corso del tempo
- Tuttavia è consentito al documento digitale di subire modifiche quali aggiunte o annotazioni autorizzate e gestite. In termini operativi, l'ente produttore può effettuare una modifica ad uno o più pacchetti di archiviazione (AIP) conservati in DigiP mediante la produzione di un nuovo SIP il cui corrispondente AIP si andrà a collegare a quello già presente nel sistema di conservazione come integrazione ovvero generando una nuova edizione (*AIP edition*) che si andrà ad affiancare alla precedente. In fase di generazione del DIP il sistema permette di visualizzare la "storia" dell'AIP e scaricarne una singola versione o l'insieme delle versioni che costituiscono la "storia" del documento

Personalizzazione delle trasformazioni

- Un’ampia parte delle trasformazioni da SIP ad AIP e da AIP a DIP è basata su XSLT (XSL Transformations) personalizzabili per ogni tipologia documentale di ogni ente. L’uso degli XSLT rende le trasformazioni customizzabili e configurabili senza bisogno di modificare il codice dell’applicativo
- Il sistema di conservazione DigiP individua nella fase di Ingest il momento in cui il SIP conferito dal soggetto produttore viene validato e quindi trasformato in AIP. Durante questo processo i risultati delle validazioni e delle conversioni di formato richieste dagli accordi di servizio e dalle politiche prestabilite vengono raccolti e aggregati in una struttura di IP (Information Package) idonea alla successiva generazione dei corrispondenti Pacchetti di archiviazione. Questa struttura transitoria identificata come KIP – Kernel Information Package - è indipendente dai formati scelti per l'archiviazione. Il file KIP è un esempio di come il processo raccoglie tutti i metadati del SIP, dei risultati delle regole e delle trasformazioni e li organizza in un file pronto per essere trasformato nel descrittore AIP

Controlli sistematici di integrità SIP-AIP

- Per ogni pacchetto versato, al termine del processo di conservazione, vengono sistematicamente effettuati una serie di controlli di integrità che garantiscono la corretta archiviazione del pacchetto stesso. Tra i controlli effettuati:
  - coerenza dell'hash degli allegati tra SIP e AIP
  - controllo della rimappatura dei metadati tra SIP e AIP

Metadati

- I metadati sono quelli previsti dalla normativa e specifici alla tipologia documentale oggetto di conservazione. Tutti i metadati trasferiti dall'ente produttore nell'indice di versamento sono mappati nell'indice dell'AIP conforme allo standard SInCRO. I metadati possono essere ricondotti alle seguenti categorie:
  - Metadati descrittivi
  - Metadati amministrativi, gestionali e di conservazione
  - Metadati strutturali
- Il sistema consente all'ente produttore **ampia flessibilità** nell'adottare metadati specifici che interessano il contesto di appartenenza del documento
- Il sistema non pone limiti al numero dei metadati associabili ad un documento
- DigiP utilizza le specifiche categorie previste dallo standard OAIS per dettagliare, distinguere e impacchettare le informazioni per la conservazione quali: descriptive information, reference information, provenance information, context information, fixity information, access information, packaging information, **accessibili all'utente da interfaccia web durante la navigazione del pacchetto di archiviazione**

Metadati di cui all'Allegato 5 delle Linee Guida sul Documento informatico

- Al fine di fornire la possibilità agli enti di versare i metadati secondo le indicazioni presenti all'Allegato 5 delle Linee Guida, ove possibile verranno utilizzati i metadati già presenti nell'interfaccia di versamento mentre ne verranno aggiunti di nuovi per tutti quei metadati ad oggi non presenti
- Inoltre al fine garantire un certo livello di interoperabilità verrà applicata una nuova trasformazione in fase di generazione del DIP che rimapperà gli attuali metadati nei nuovi metadati previsti dall'Allegato 5

Criticità: mancanza di requisiti tecnici per consentire la piena interoperabilità tra i diversi poli

Sistema documentale per agevolare l’ente produttore nel trattamento dei documenti conservati

- Gli utenti, autorizzati ed abilitati, possono accedere al sistema di conservazione DigiP per monitorare il processo di conservazione e ricercare i documenti digitali conservati. L’utente attualmente dispone di due aree funzionali: **“Ingest”** attraverso la quale può effettuare il versamento di un pacchetto di versamento (SIP), consultare la lista dei SIP ricevuti dal sistema, visualizzare ed effettuare il download del Rapporto di versamento, dei SIP e dei log applicativi; **“Access”** attraverso la quale può ricercare i documenti conservati ed effettuarne il download tramite la generazione del DIP (pacchetto di distribuzione)
- Il DIP scaricato può essere di due tipi: singolo (corrispondente al singolo AIP selezionato) o Completo (corrispondente all’AIP selezionato e a tutti i suoi aggiornamenti succedutisi nel tempo)

Autorizzazione e autenticazione

- DigiP mette a disposizione un meccanismo di autorizzazione basato su Access Control List e un sistema di autenticazione integrato con il Portale Servizi di Regione Marche che permette già all'utente di autenticarsi tramite SPID

Policy di visibilità dei documenti in base al modello organizzativo

- DigiP permette la configurazione di policy di visibilità dei documenti, non solo legate al livello di riservatezza, ma anche al modello organizzativo del soggetto produttore, consentendo **visibilità** ad esempio limitate **a specifiche Unità Organizzative del soggetto produttore**

Prospettive future per i Poli di conservazione
Qualità: una caratteristica irrinunciabile

- Nel tempo, molti sono stati i passi avanti fatti, ma molto rimane ancora da fare per la qualità dei nostri archivi digitali e per sensibilizzazione della cultura della qualità degli archivi nei soggetti produttori.
- Il Polo di conservazione Marche Digip intende continuare ad investire su questo fronte come ha fatto negli ultimi anni, collaborando con i professionisti del settore, con controlli automatici di coerenza e consistenza, con controlli a campione sulla qualità dei dati forniti dagli enti e con un confronto diretto con questi ultimi

Prospettive future per i Poli di conservazione
Interoperabilità: un requisito essenziale

- Accanto alla normativa già esistente, agli standard di settore e ovviamente alla cultura archivistica ora anche le nuove linee guida Agid contribuiscono a definire sempre meglio le caratteristiche dei nostri archivi digitali,
- Tuttavia, dall'analisi dei Pacchetti Informativi gestiti dai vari conservatori, ci si rende conto che le differenze di forma e di sostanza sono ancora molte
- Occorre quindi lavorare sulla interoperabilità, quanto meno per i Poli pubblici, per garantire che ai nostri archivi non solo l'integrità, ma anche continuità nella gestione e nella rappresentazione semantica.
- L'interoperabilità non ci garantirà solo la possibilità di far gestire nel tempo un archivio a conservatori diversi, ma anche la possibilità di implementare sistemi di lettura integrata di più archivi

Prospettive future per i Poli di conservazione
Big Data: una possibilità da sfruttare

- Gli archivi sono diventati digitali, questo significa da una parte riduzione di cosi e spazio per l'archiviazione, dall'altra difficoltà di gestione di formati e obsolescenza tecnologica, ma significa anche opportunità.
- Sfruttando le nuove tecnologia si aprono, anche per gli archivi possibilità finora inesplorate come l'analisi dei documenti attraverso i big data o la possibilità di ricercare contenuti sistemi automatici resi intelligenti e funzionali dal machine learning

Grazie

https://www.regione.marche.it/Regione-Utile/Agenda-Digitale/Polo-di-conservazione-regionale