

Testo tratto da S. PIGLIAPOCO, *La memoria digitale delle amministrazioni pubbliche. Requisiti, metodi e sistemi per la produzione, archiviazione e conservazione dei documenti informatici*, Santarcangelo di Romagna (RN), Maggioli editore, 2005

Il modello OAIS

Il Reference Model for an Open Archival Information System (OAIS) – standard ISO 14721:2003 - è un modello concettuale di sistema informativo per l'archiviazione e la conservazione delle risorse digitali. È stato elaborato dal CCSDS (Consultative Committee for Space Data System)¹ e rappresenta uno standard ISO (International Organization for Standardization) in fase di elaborazione con il contributo di enti ed istituzioni diverse. Esso si autodefinisce «un'organizzazione di soggetti e sistemi che hanno accettato la responsabilità della conservazione dell'informazione e del mantenerla disponibile per una comunità designata».

Gli elementi che maggiormente caratterizzano il modello OAIS sono:

- La natura di *soluzione aperta*, nel senso che si presenta come standard ISO, aperto, disponibile alla comunità internazionale e ottimizzabile con il contributo di tutti;
- l'*applicabilità a qualsiasi oggetto informativo digitale*, che viene visto come sequenza di bit associata ad un insieme di informazioni che ne permettono la rappresentazione e la comprensione a livello utente;
- l'essere personalizzato per una *comunità designata*, che presuppone lo studio preliminare delle esigenze del soggetto produttore e degli utenti del sistema;
- la *concretezza*, che discende dall'essere disegnato per una comunità ben identificata con la quale sono stati concordati preventivamente tutti gli aspetti connessi al conservazione e fruizione del patrimonio informativo digitale;
- la *responsabilità della conservazione digitale*, nel senso che presuppone l'esistenza di una struttura che definisca i metodi per la conservazione delle risorse digitali e li attui;
- la possibilità di sviluppare *sistemi di archiviazione distribuiti e interoperabili*, che nasce dalla capacità tecnica di far "parlare" tra di loro due o più sistemi OAIS, ciascuno personalizzato per una certa comunità, oppure di includere in un sistema OAIS componenti aggiuntivi che gli permettono di servire contemporaneamente più comunità;

¹ Il CCSDS è nato per volontà delle maggiori agenzie spaziali del mondo al fine di elaborare raccomandazioni e specifiche tecniche per la condivisione di dati. Informazioni dettagliate sul CCSDS e sul modello OAIS sono reperibili sul sito www.ccsds.org

- la possibilità di applicare *metodi di conservazione diversi*, anche se viene privilegiato il metodo basato sulla migrazione rispetto a quello basato sull'emulazione².

I soggetti che interagiscono con un sistema OAIS sono rappresentati in figura 2.1.

I *Produttori* sono le persone fisiche o giuridiche, pubbliche o private, che producono gli oggetti informativi da conservare nel sistema OAIS, rispettando i requisiti e le procedure preventivamente concordate con il Management e il responsabile dell'Archivio; i *Consumatori* sono i soggetti interessati all'accesso e alla fruizione del patrimonio digitale conservato nel sistema OAIS; il *Management* è la struttura a cui compete la responsabilità di alto livello, ossia la definizione delle politiche e degli obiettivi generali dell'OAIS; l'*Archivio* è la struttura responsabile della trattazione quotidiana degli oggetti informativi archiviati, con procedure e metodi che ne assicurano la conservazione e fruizione a lungo termine.

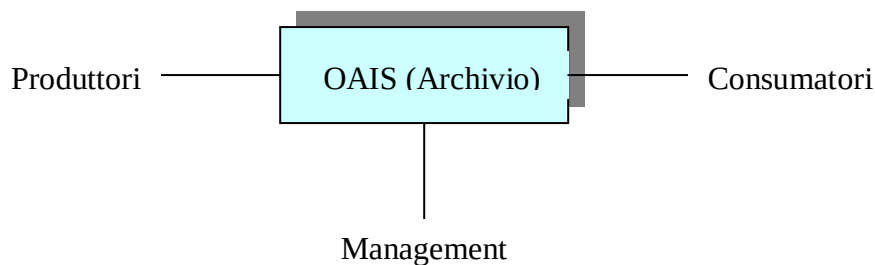


Figura 2.1 - Ambiente operativo di un sistema OAIS

Per descrivere il funzionamento di un sistema OAIS occorre prima chiarire il concetto di *pacchetto informativo* (information package) che è alla base del modello sviluppato dal CCSDS. Esso si compone di quattro parti: le *informazioni di contenuto* (content information), le *informazioni di conservazione* (preservation description information), le *informazioni di pacchetto* (packaging information) e i *dati descrittivi del pacchetto* (package description). In figura 2.2 è riportato lo schema sintetico di un information package.

² Vedi il precedente paragrafo 1.5.3

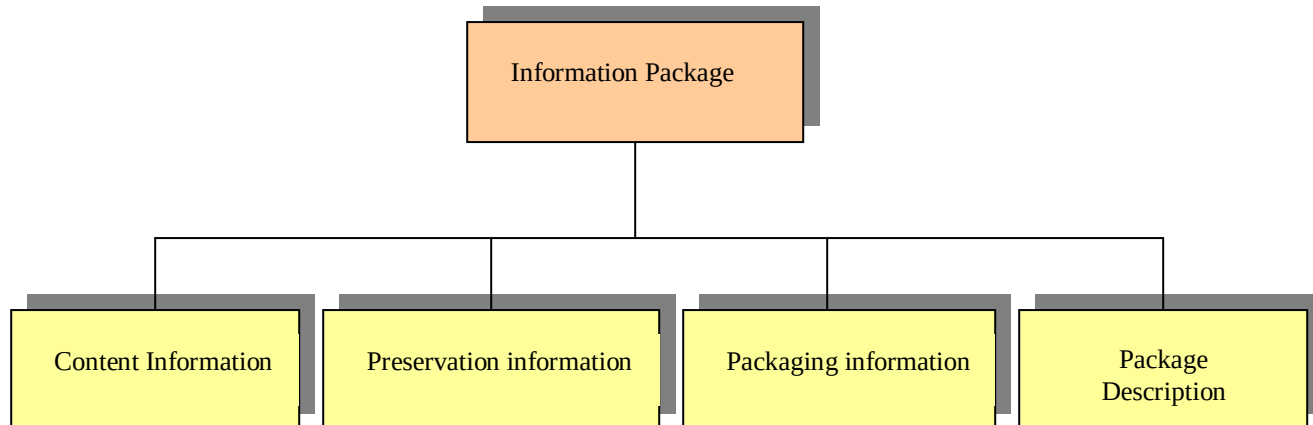


Figura 2.2 - Schema sintetico di un pacchetto informativo OAIS

Le informazioni di contenuto comprendono l'*oggetto informativo* (data object), che può essere digitale (digital object) oppure fisico (physical object)³, e l'insieme delle informazioni che ne permettono la rappresentazione e la comprensione a livello utente (*representation information*). Nel caso di oggetti informativi digitali, le informazioni di rappresentazione ne specificano il formato, ovvero l'insieme dei codici e delle regole che, a partire dalla sequenza binaria che costituisce il digital object, permettono di riprodurre l'oggetto stesso, a video o a stampa, nella forma e nel contenuto originario⁴. Vista la complessità delle representation information, soprattutto in relazione ad entità digitali che contengono immagini, filmati o registrazioni sonore, è prevista la possibilità di organizzarle a catena (*representation network*), nel senso che all'interno di un insieme di informazioni di rappresentazione è possibile richiamarne altre di livello più specifico e così via fino a raggiungere quella di base.

Le informazioni di conservazione specificano i metadati⁵ che identificano l'oggetto informativo, lo qualificano sotto il profilo dell'integrità e lo collocano nel contesto di

³ Si noti come il modello OAIS possa essere utilizzato per la conservazione di un archivio composto in parte da documenti informatici e in parte da documenti cartacei in quanto l'organizzazione dei metadati è la stessa sia per gli uni che gli altri. Si sottolinea, inoltre, che gli oggetti informativi digitali sono trattati come sequenze di bit, ciascuno dei quali è un'entità a sé stante.

⁴ Le representation information contengono informazioni sia di natura sintattica che semantica e svolgono un ruolo analogo a quello che il dizionario delle lingua italiana e le regole grammaticali svolgono per la comprensione di un testo scritto in italiano.

⁵ I metadati possono essere definiti come "dati sui dati". Il concetto viene dall'informatica e precisamente dalla teoria sulla banche dati dove vengono utilizzati per descrivere i gruppi di dati memorizzati nel database. Peter Horsman, in suo studio sul concetto archivistico di metadata, ne propone una definizione più puntuale: "i metadata di archiviazione sono informazioni strutturate e semi-strutturate che permettono la creazione, la gestione e l'uso di documenti archivistici attraverso un certo tempo e un certo ambito. I metadata di archiviazione possono identificare, autenticare e contestualizzare i documenti archivistici e le persone, i processi e i sistemi che li creano, li gestiscono e li usano" [cfr. P.

provenienza. Esse comprendono le *informazioni di identificazione* (reference information), le *informazioni di contesto* (context information), le *informazioni di provenienza* (provenance information) e le *informazioni di stabilità* (fixed information). Le informazioni di identificazione forniscono i dati attraverso i quali l'OAIS potrà individuarlo all'interno del patrimonio digitale conservato; esse riguardano l'oggetto informativo visto nel suo complesso, ma possono riferirsi anche alle sue singole parti costitutive⁶. Le informazioni di contesto descrivono l'ambiente tecnologico nel quale è stato prodotto il data object, con particolare riferimento ad eventuali link di natura informatica che lo collegano ad altri data object, alle eventuali migrazioni che lo hanno interessato e al livello di sicurezza dell'impianto tecnologico da cui ha avuto origine. Le informazioni di provenienza esplicitano il contesto organizzativo e istituzionale nell'ambito del quale è stato prodotto l'oggetto informativo, evidenziando le vicende che lo hanno interessato dalla nascita fino al versamento nel sistema OAIS e, soprattutto, le responsabilità connesse alla sua custodia. Infine, le informazioni di stabilità forniscono gli strumenti per verificare l'integrità dal data object tra cui: la firma elettronica, la marca temporale, gli audit trail, le evidenze informatiche che permettono di rilevare le anomalie nella memorizzazione, lettura o trasmissione dei dati binari (codici CRC - Cyclic Redundancy Check).

Le packaging information specificano i dati che indirizzano alle posizioni delle componenti del pacchetto informativo nel sistema di storage. Ad esempio, se le informazioni di contenuto e le informazioni di conservazione relative ad un data object sono memorizzate in un file registrato su CD-ROM, le informazioni di pacchetto riportano le indicazioni sul nome del file e sul percorso da seguire per raggiungerlo (directory). Questo insieme di dati permettono di accedere a tutte le versioni di un data object a partire da quella di origine versata del soggetto produttore e proseguendo attraverso le successive, generate con processi di migrazione, fino ad arrivare a quella corrente.

I dati informativi sul pacchetto (package description) hanno lo scopo di agevolare l'utente nella ricerca e nell'acquisizione degli oggetti informativi conservati nel sistema OAIS. Normalmente, essi derivano dalle informazioni di contenuto e dalle informazioni di conservazione e realizzano tutti i possibili punti di accesso al pacchetto informativo, visto sia come entità indivisibile che come insieme di parti disgiunte accessibili separatamente.

In figura 2.3 è riportato lo schema che rappresenta in modo analitico un information package⁷.

HORSMAN, *Metadata: concetto archivistico o territorio informatico*, in *Archivi & Computer*, Anno XI, n. 1, S. Miniato (PI), 2001]

⁶ Ad esempio, ciò è possibile se si utilizza il linguaggio XML per identificare le parti costitutive di un documento informatico.

⁷ Gli schemi riportati nelle figure 2.3, 2.4, 2.5, 2.6, e 2.7 sono tratti da CONSULTATIVE COMMITTEE FOR SPACE DATA SYSTEMS, *Reference model for an Open Archival Information System (OAIS)*, CCSDS 650.0-B-1, Blue book, 2002

Si comprende già da queste note che il modello OAIS mira alla conservazione delle risorse digitali non solo come oggetti a sé stanti, ma come entità collegate tra di loro e all'ente produttore, alla sua organizzazione, al suo impianto tecnologico, al contesto politico e istituzionale in cui esso opera.

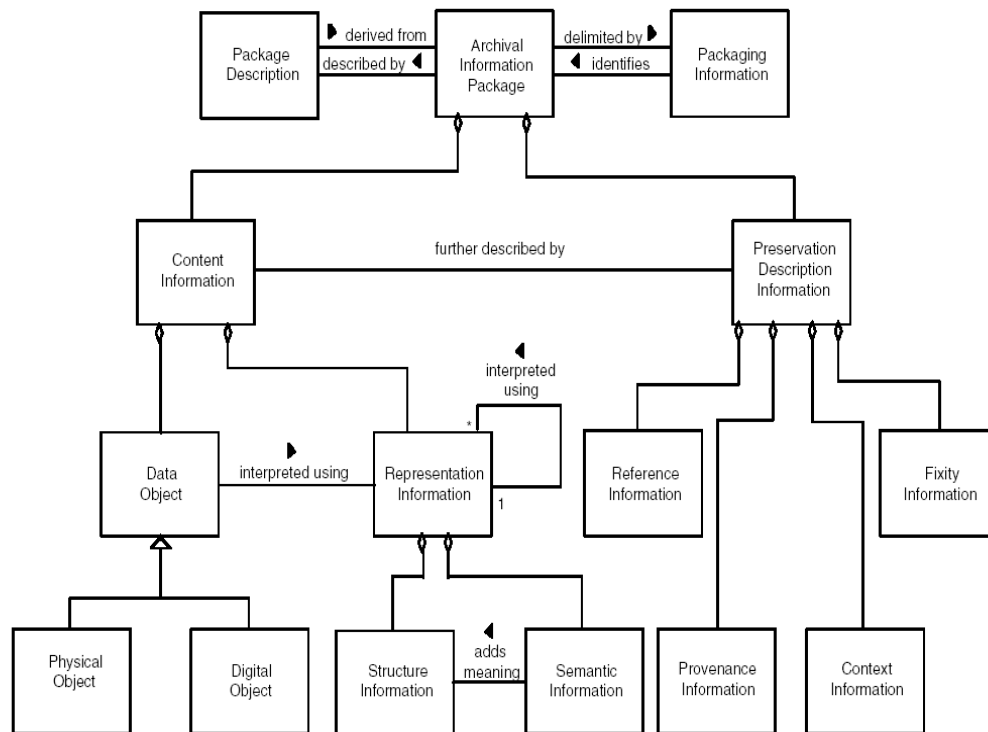


Figura 2.3 - Schema analitico di un pacchetto informativo OAIS

Un soggetto produttore può immettere risorse digitali da conservare nel sistema OAIS inviandole sotto forma di pacchetti informativi, che prendono il nome di *Submission Information Packages* (SIP), associati ai metadati sopra descritti. Tali pacchetti sono verificati dal sistema con un'apposita funzione ed integrati con le informazioni mancanti, dando origine ai pacchetti destinati alla conservazione a lungo termine che prendono il nome di *Archival Information Packages* (AIP). Inoltre, sulla base degli accordi stipulati con il soggetto produttore, il sistema OAIS può dare origine ad altri pacchetti informativi, che prendono il nome di *Dissemination Information Packages* (DIP), allo scopo di soddisfare al meglio le esigenze dei consumatori relativamente alla ricerca e all'acquisizione delle risorse digitali di loro interesse. In figura 2.4 è riportato lo schema dei flussi dei pacchetti informativi tra il soggetto produttore, il sistema OAIS e i consumatori.

Un Archival Information Package può essere trattato come un oggetto a sé stante, e in tal caso viene denominato *Archival Information Unit* (AIU), oppure come parte costitutiva di un insieme di pacchetti informativi legati tra di loro da una qualche relazione, che prende il nome di *Archival Information Collection* (AIC). Naturalmente, una AIC può essere a sua volta collegato ad altri AIC e così via. L'Archival Information Collection è un'entità che assomiglia molto a quelle che nel paragrafo 1.4 abbiamo chiamato fascicolo e serie di un fondo archivistico; ad essa deve essere associato l'insieme dei metadati che la descrivono e che permettono agli utenti di accedervi per consultazione o acquisizione.

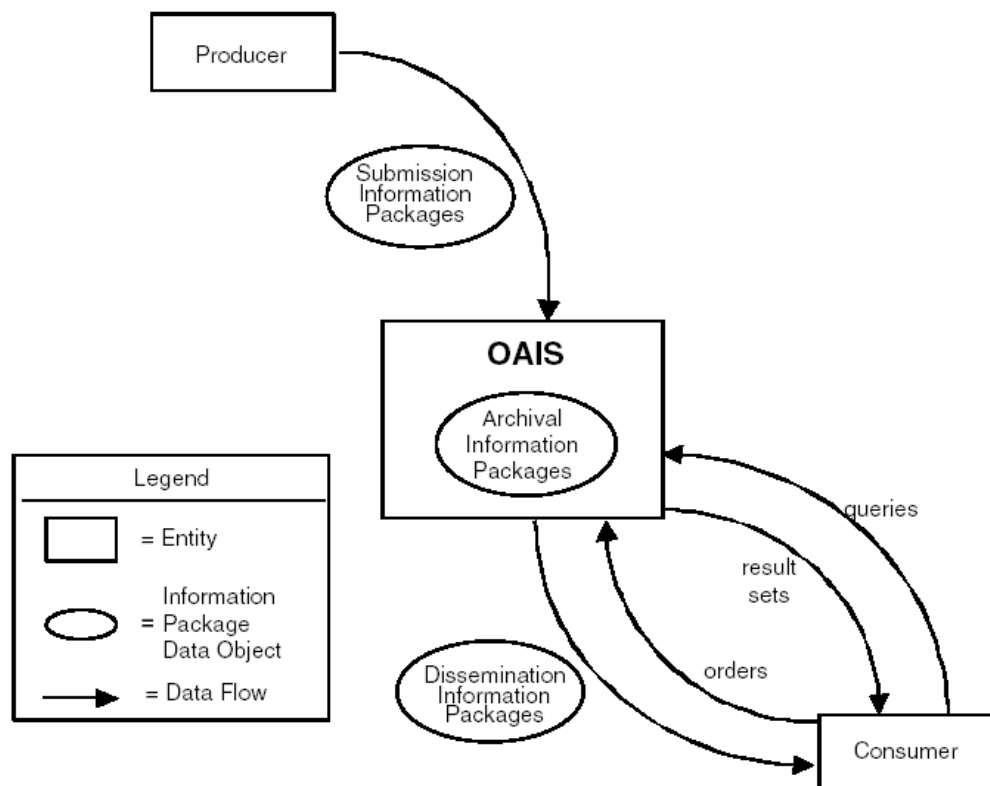


Figura 2.4 - Schema dei flussi dei pacchetti informativi da e verso l'OAS

Le funzionalità di un sistema OAS sono raggruppate in sei moduli: INGEST; DATA MANAGEMENT; ARCHIVAL STORAGE; ADMINISTRATION; PRESERVATION PLANNING; ACCESS. In figura 2.5 è riportato il modello funzionale di un sistema OAS.

Il modulo di INGEST si occupa della:

- ricezione dei pacchetti informativi (SIP) trasmessi dall'ente produttore per essere immessi nel sistema OAS (receive submission);
- verifica dei SIP al fine di accertarne l'integrità e la completezza (quality assurance);

- emissione di una ricevuta di presa in carico che determina l'assunzione di responsabilità da parte della struttura di conservazione;
- generazione degli Archival Information Packages (generate AIP);
- trasmissione degli AIP generati al modulo di ARCHIVAL STORAGE per la memorizzazione su supporti fisici.

Il modulo di DATA MANAGEMENT utilizza un database management system per gestire i dati sui pacchetti informativi archiviati e sul funzionamento del sistema OAIS.

Il modulo di ARCHIVAL STORAGE riceve i pacchetti AIP dal modulo di INGEST e li memorizza sui media del sistema di storage che saranno scelti in base al periodo di conservazione stimato per l'AIP in fase di generazione. Tale modulo comprende anche le funzionalità per il controllo degli errori in lettura e scrittura sui media (error checking), il trasferimento degli AIP da un media ad un altro (replace media), l'esecuzione delle operazioni di backup e disaster recovery.

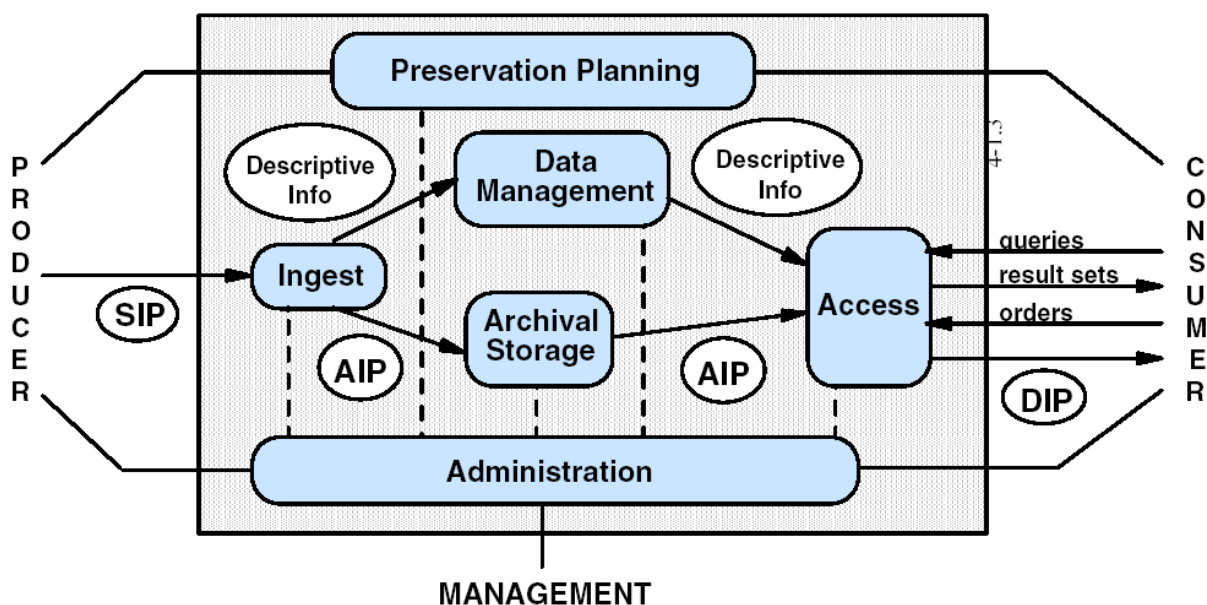


Figura 2.5 - Modello funzionale di un sistema OAIS

Il modulo di ACCESS realizza un'interfaccia attraverso la quale i consumatori possono inoltrare al sistema OAIS le loro richieste di accesso. In linea generale, sono previste tre possibilità: a) consultazione di uno o più pacchetti informativi archiviati; b) richiesta di report di varia natura; c) richiesta di acquisizione di una o più entità conservate. In quest'ultimo caso, a partire dall'AIP richiesto, potranno essere prodotti uno o più Dissemination Information Package (DIP) da rendere disponibili ai soggetti interessati.

Il modulo di ADMINISTRATION realizza:

- la gestione della configurazione dei pacchetti informativi (SIP e AIP) in base agli accordi stipulati con l'ente produttore (negotiated submission agreement e audit submission);
- la gestione della configurazione del sistema OAIS (manage system configuration) attraverso il monitoraggio continuo della sua operatività (audit trail), la registrazione di ogni modifica apportata alla configurazione, il controllo dell'integrità del sistema per l'intero ciclo di vita (gestione delle migrazioni del sistema);
- i meccanismi di autenticazione e autorizzazione degli utenti, nonché le misure di protezione volte ad evitare accessi non autorizzati alle risorse informative conservate;
- la gestione degli standard e delle regole (establish standards and policies) che riguarda: a) la scelta degli standard di formato e di documentazione; b) la definizione delle procedure di verifica per l'accettazione dei SIP; c) le regole per la migrazione dei pacchetti informativi e la sostituzione dei media obsoleti; d) la definizione delle procedure di backup e disaster recovery;
- la gestione dei servizi ai clienti (customer service) che include: a) la creazione, l'aggiornamento e la cancellazione degli account; b) l'emissione delle fatture; c) la erogazione di servizi informativi.

Il modulo di PRESERVATION PLANNING si occupa:

- del monitoraggio della comunità designata (monitor designed community) al fine di acquisire informazioni su nuove esigenze di servizio;
- del monitoraggio della tecnologia (monitor technology) per individuare quella emergente e i nuovi standard;
- della definizione della strategia di sviluppo del sistema (develop preservation strategy and standards) in modo che si possa accogliere i nuovi standard tecnologici e soddisfare al meglio le esigenze della comunità designata;
- dello sviluppo di nuovi modelli di pacchetto informativo e di piani dettagliati per la migrazione digitale (develop packaging designs and migration plan).

I sistemi OAIS permettono di realizzare diverse soluzioni per la conservazione delle risorse digitali. Quella più semplice è denominata *Independent Archives* e vede la presenza di un sistema OAIS che offre servizi ad una comunità designata senza preoccuparsi di interagire e scambiare informazioni con altri sistemi. Al livello immediatamente superiore c'è la soluzione denominata *Cooperating Archives*, che si basa sull'interoperabilità tra due o più sistemi OAIS realizzata attraverso la condivisione dei modelli di SIP e DIP. In un'architettura di questo genere un sistema A può inviare una richiesta di materiale informativo ad un altro sistema B,

ottenendo come risposta il pacchetto cercato⁸. Una terza soluzione è denominata *Federated Archives*, che realizza un modello architetturale in cui più sistemi OAIS indipendenti condividono una stessa interfaccia utente per consentire ad una comunità globale di consumatori di accedere a più archivi con le stesse modalità. Tale soluzione architetturale richiede la costituzione di un catalogo comune del patrimonio informativo complessivamente reso disponibile e lo smistamento delle richieste degli utenti ai sistemi OAIS che sono depositari delle informazioni cercate. Essa è rappresentata in figura 2.6. Una quarta soluzione è denominata *Shared Resources Archives* e vede un sistema OAIS servire più enti produttori diversi e più comunità di consumatori, grazie all'attivazione di componenti aggiuntive specializzate. In altri termini, un sistema OAIS utilizzato in architettura shared resources si presenta con i moduli di ARCHIVAL STORAGE e DATA MANAGEMENT condivisi e gli altri duplicati in quantità sufficiente per servire le diverse comunità che si avvalgono dei suoi servizi. Tale architettura è rappresentata in figura 2.7. Infine, appare possibile combinare un'architettura di tipo Federated Archives con una di tipo Shared Resources Archives per permettere ad una comunità globale di utenti di accedere a più sistemi OAIS ciascuno dei quali è utilizzato da più soggetti produttori.

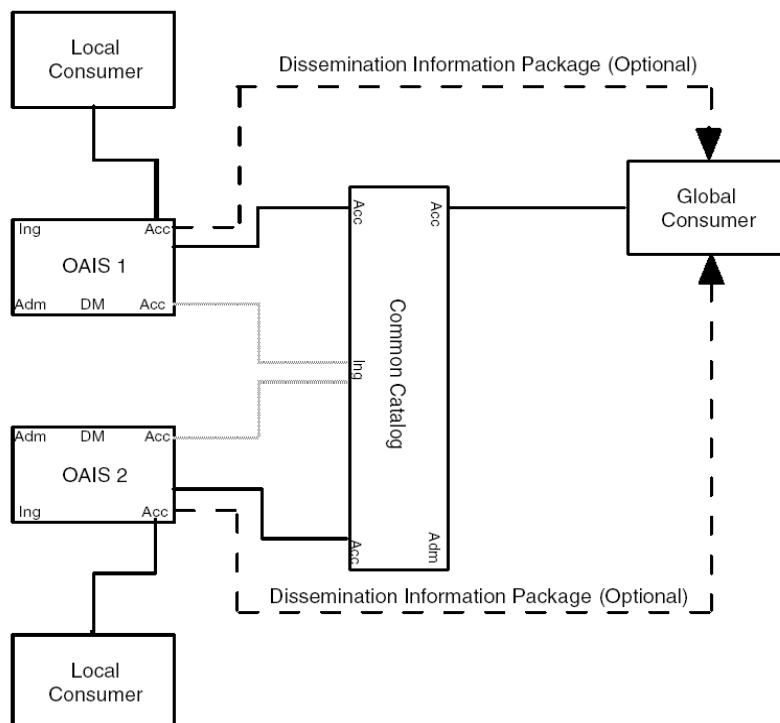


Figura 2.6 - Schema architetturale di tipo Federated Archives

⁸ In questo caso il pacchetto informativo rilasciato dal sistema B (DIP) a fronte della richiesta inoltrata dal sistema A viene trasmesso a quest'ultimo sotto forma di pacchetto informativo d'ingresso (SIP). Pertanto, i due pacchetti informativi, il DIP del sistema B e il SIP del sistema A, devono esser compatibili tra di loro.

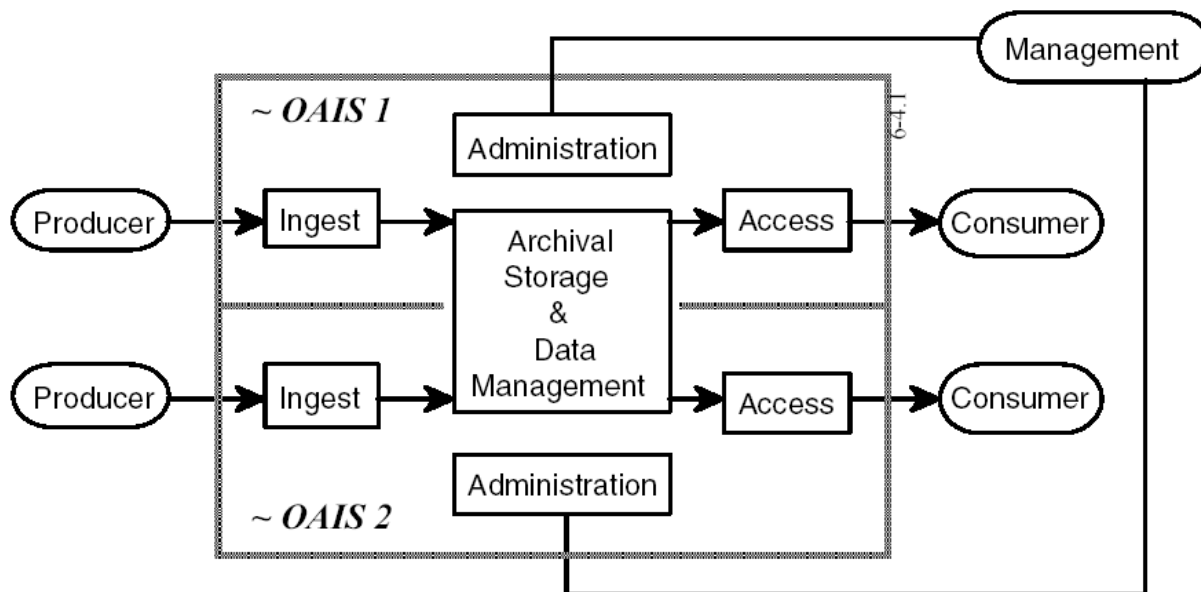


Figura 2.7 - Schema architetturale di tipo Shared Resources Archives