

Dig *Italia*

Anno III, Numero 1 - **2008**

Rivista del digitale nei beni culturali

ICCU-ROMA

Il modello OAIS*

Giovanni Michetti

Università di Roma "La Sapienza"

Nel corso degli ultimi anni si sono moltiplicate le iniziative volte ad individuare soluzioni per la conservazione digitale a lungo termine, alcune mirate alla definizione di strutture di metadati per gestire gli elementi informativi necessari a sostenere i processi conservativi; altre dedicate allo studio delle tecniche di trasferimento dell'informazione da un contesto obsoleto verso nuovi supporti, sistemi e ambienti; altre ancora orientate all'analisi dei rapporti costi-benefici correlati alle diverse strategie conservative; non ultime, alcune iniziative hanno approfondito l'analisi teoretica con l'obiettivo di identificare le componenti significative degli oggetti documentali. In realtà, il problema della conservazione digitale è così complesso e urgente che occorre agire contemporaneamente in tutte le direzioni, muovendo dalla dimensione tecnica a quella teoretica, senza trascurare il profilo organizzativo e quello socio-culturale, giacché la costruzione di una politica della conservazione passa inevitabilmente per la sensibilizzazione e il coinvolgimento di soggetti della società non racchiusi nei confini ristretti della comunità di dominio.

Il modello di riferimento OAIS (Open Archival Information System) si propone come quadro concettuale unitario per descrivere oggetti, processi, strategie e tecniche finalizzati alla conservazione digitale a lungo termine, nonché per comprendere le loro reciproche relazioni e per analizzare e confrontare soluzioni conservative diverse. Il modello si basa su un'idea dinamica della conservazione, intesa come processo permanente e mai concluso di monitoraggio del contesto in cui sono immersi gli oggetti; e disegna una complessa architettura di funzioni e oggetti informativi fondata sull'individuazione delle risorse necessarie per ricostruire il significato degli oggetti, assumendo il bit come unità minima del sistema concettuale.

Introduzione

La funzione conservativa in ambiente digitale è uno temi di maggior impegno per la comunità scientifica di settore: l'innovazione tecnologica ha infatti determinato una vera e propria rivoluzione nei processi di comunicazione e – di conseguenza – nei sistemi documentari contemporanei, chiamati a confrontarsi con nuovi oggetti, metodi e politiche necessari per garantire la continuità della funzione documentale all'interno di un contesto completamente rinnovato. Tuttavia, mentre in materia di formazione e gestione delle risorse digitali disponiamo ormai

* Pubblichiamo in queste pagine, con alcune modifiche e integrazioni, il saggio introduttivo al volume *OAIS. Sistema informativo aperto per l'archiviazione*, a cura di Giovanni Michetti, pubblicato dall'ICCU nella collana *La conservazione del digitale*.

di una consolidata esperienza e di un'ampia letteratura, entrambe assunte a fondamento per l'elaborazione di norme, linee guida e modelli a carattere nazionale e internazionale, non altrettanto può dirsi per il tema della conservazione digitale, caratterizzato ancora da un'incertezza generale in merito a soluzioni e strategie. Tale fenomeno non sorprende, poiché lo sviluppo di conoscenze e strumenti specifici per affrontare adeguatamente il problema conservativo avviene in contrasto con la natura stessa degli oggetti documentali contemporanei, caratterizzati da un'estrema instabilità, sia in relazione al profilo fisico sia con riferimento al contesto logico di appartenenza: il degrado dei supporti materiali e l'obsolescenza dei dispositivi di scrittura e memorizzazione costituiscono l'aspetto più evidente del processo di degrado della memoria documentaria, ma non meno rilevante è la perdita del sistema di relazioni nel quale un documento è inserito. Il progresso tecnologico ha infatti condotto alla dematerializzazione dei documenti, ne ha cioè ridotto la fisicità trasformandoli in entità non più autonome bensì fortemente dipendenti da un insieme di collegamenti, e oscurando quei vincoli logici – altrimenti evidenti in ambito tradizionale – che conferiscono significato agli oggetti documentali: sempre più spesso la rappresentazione di un documento non è altro che il risultato dell'aggregazione istantanea di componenti digitali distinte all'origine, come banalmente esemplificato da una qualunque pagina Web; e sempre più spesso i sistemi documentari digitali sono purtroppo strutturati in maniera drammaticamente puntiforme, immergendo i documenti all'interno di depositi pressoché privi di organizzazione e dimenticando che gran parte del significato degli oggetti risiede proprio nelle relazioni cui partecipano, come risulta evidente – ad esempio – dal ruolo essenziale che l'organizzazione gerarchica di un archivio assume per la comprensione delle sue carte.

Inoltre, la missione conservativa trova un ulteriore e paradossale limite nei suoi stessi obiettivi: l'esigenza di mantenere l'integrità dei documenti digitali deve essere confrontata e mediata con la necessità di garantirne l'accessibilità nel tempo. Ciò significa, nella sostanza, bilanciare la tendenza a fissare e consolidare gli oggetti documentali e il loro contesto di riferimento, con gli inderogabili e continui interventi di copiatura, conversione e migrazione richiesti per fronteggiare l'obsolescenza dei sistemi documentari contemporanei: un equilibrio ancora tutto da scoprire fra la staticità della cristallizzazione degli oggetti e la dinamicità delle modifiche imposte dalla loro natura.

Conservare in ambiente digitale, dunque, costituisce la sfida del futuro, un'attività complessa che non può essere ridotta al solo profilo materiale, ma che deve volgere lo sguardo al sistema di relazioni logiche interne ed esterne ai documenti, per coglierne i livelli di significato ritenuti qualitativamente rilevanti e meritevoli di preservazione: occorre quindi lo sviluppo di una solida base concettuale e metodologica, al fine di elaborare un modello complessivo in grado di garantire la conservazione permanente dei documenti.

In questo contesto, il modello OAIS (Open Archival Information System) rappresenta una seria proposta di analisi e descrizione degli oggetti e delle dinamiche della *digital preservation*, un tentativo di sistematizzazione teoretica, un quadro concettuale che non ha mancato di suscitare l'attenzione della comunità scientifica in ragione dei contenuti innovativi e delle rigorose metodologie formali. Tali caratteristiche giustificano l'elevata coerenza interna del disegno complessivo, mentre l'originalità della prospettiva di analisi costringe a ripensare gli oggetti e i processi della funzione conservativa: l'esistenza di una prassi e di una dottrina consolidata in materia non consente di eludere la necessità di adottare logiche nuove che, seppur radicate nel *corpus* di conoscenze di dominio, assumono ormai un autonomo rilievo. La legittima aspirazione a riconoscere le omogeneità e rintracciare un filo conduttore che, dalla carta al bit, consenta di interpretare i fenomeni in maniera unitaria deve cedere il passo alle evidenti peculiarità dell'ambiente digitale, in misura tale da non rendere possibile un approccio senza soluzione di continuità.

Il modello OAIS assume quindi un valore privilegiato in quanto strumento per leggere la realtà degli oggetti digitali e affrontarne la complessità con un approccio unitario: non si limita infatti a definire quella che oggi a buon diritto definiremmo un'ontologia per la conservazione, ma fornisce anche una guida per individuare delle strategie conservative in relazione tanto al profilo tecnico quanto a quello organizzativo. E propone un'architettura di sistema che non solo possa essere assunta a fondamento per la progettazione di moduli e servizi dedicati alla conservazione, ma che costituisca al contempo un terreno per la riflessione teorica sui nodi critici: la sistematizzazione dei principi fondanti per la conservazione in ambiente digitale non è operazione neutrale e asettica che possa compiersi senza precisi obiettivi e chiavi di interpretazione afferenti a uno specifico *milieu*. Il modello incarna tali caratteristiche e come tale può rappresentare un fecondo luogo di confronto e valutazione di soluzioni differenti, uno strumento di mediazione utilizzabile in contesti disciplinari affini ma diversi, un efficace veicolo di promozione di una cultura condivisa della conservazione digitale, con la consapevolezza dell'importanza di consolidare linguaggi, tecniche e teorie al fine di configurare stabilmente una comunità italiana dedicata al tema della *digital preservation*.

Genesi

Il Consultative Committee for Space Data Systems (CCSDS – Comitato consultivo per i sistemi di dati spaziali) nasce formalmente nel 1982 come organo di coordinamento per le agenzie spaziali interessate allo sviluppo cooperativo di standard per la gestione dei dati relativi alla ricerca spaziale: attraverso le analisi di settore e lo sforzo congiunto di esperti internazionali, il comitato promuove la ricerca di soluzioni condivise che, elaborate nella forma di raccomandazioni, sono riconosciute e adottate da tutti i membri dell'organismo. A partire dal 1990 il CCSDS avvia un

accordo di cooperazione con l'International Standard Organization (ISO) al fine di sottoporre le proprie raccomandazioni alle procedure formali di revisione e approvazione necessarie a garantirne il riconoscimento di standard internazionali: su richiesta dell'ISO, il CCSDS assume il coordinamento dello sviluppo di standard per l'archiviazione a lungo termine dei dati digitali generati dalle osservazioni geospaziali. Ben presto, però, le attività di ricerca evidenziano l'assenza di un quadro concettuale e metodologico sufficientemente stabile e condiviso, tale da poter essere assunto a fondamento per la costruzione di standard di settore; e pertanto il CCSDS assume come prioritaria l'esigenza di elaborare un modello di riferimento per un sistema informativo aperto per l'archiviazione: è l'atto di nascita del *Reference Model for an Open Archival Information System*, elaborato in forma di bozza nel 1997 e successivamente approvato prima come raccomandazione ufficiale (Blue Book) e poi come standard ISO 14721:2003¹.

Caratteristiche generali

Un sistema informativo aperto per l'archiviazione (OAIS) è

«un archivio, inteso come struttura organizzata di persone e sistemi, che accetta la responsabilità di conservare l'informazione e renderla disponibile per una comunità di riferimento».

Questa definizione evidenzia molto chiaramente il ruolo primario che la conservazione e l'accesso assumono nella complessa architettura proposta dal modello, ove il termine aperto si riferisce al fatto che il processo di elaborazione dello standard si è svolto condividendo in ampia misura le conoscenze e utilizzando il contributo di tutti i soggetti interessati. La definizione di OAIS è apparentemente adeguata per descrivere un qualunque istituto di conservazione, sia esso archivio, biblioteca o altro ente, ma in realtà è strettamente vincolata non solo a un apparato definitorio che – come vedremo più avanti – chiarisce e circoscrive il significato di comunità di riferimento, ma anche a precise responsabilità puntualmente enunciate nel documento:

- concordare con i produttori di informazioni le risorse informative da sottoporre ad archiviazione, stabilendo criteri certi di acquisizione e selezione, e delineando così i confini stessi dell'OAIS;
- ottenere un controllo delle risorse archiviate sufficiente a garantire la conservazione a lungo termine;
- determinare la comunità di riferimento, individuandone caratteristiche, natura ed esigenze;

¹ Il documento, nella forma di raccomandazione del CCSDS, è disponibile all'indirizzo <http://public.ccsds.org/publications/archive/650x0b1.pdf>.

- assicurare che l'informazione conservata sia comprensibile autonomamente dalla comunità degli utenti, senza che si debba fare ricorso all'intervento dei soggetti che hanno prodotto l'informazione;
- adottare delle politiche e delle procedure che forniscano delle ragionevoli garanzie di sicurezza in materia di conservazione e che consentano la diffusione di copie riconoscibili come autentiche o di cui sia comunque chiara la relazione con l'originale;
- rendere disponibile all'utenza le risorse dell'archivio.

La corretta comprensione di tali requisiti non può tuttavia prescindere dalla lettura dell'articolato sistema di definizioni sul quale il modello OAIS è basato, e che risponde alla fondamentale esigenza di costruire uno strumento di mediazione per ambiti disciplinari differenti: di qui l'assenza di una specifica terminologia di dominio, nonché la rielaborazione di concetti apparentemente scontati. In questo senso deve essere letta, ad esempio, l'adozione del termine *ingest* per identificare ciò che i bibliotecari chiamano accessione e gli archivisti acquisizione, ma che nella sostanza coincide in entrambi i casi con l'ingresso di una risorsa informativa nel sistema documentario; ancor più interessante la rielaborazione di taluni concetti, come quello di conservazione a lungo termine, ove lo sguardo non è rivolto come al solito a un indefinito futuro, a un intervallo temporale così ampio da nascondere il suo limite superiore, bensì a un arco cronologico sufficientemente ampio da essere coinvolto nei fenomeni di obsolescenza tecnologica e di evoluzione della comunità di utenti. Il lungo termine non è definito *per negativum*: il modello OAIS sembra dirci che è qui e ora, e che non è altro se non una serie infinita di "piccoli termini". Si noti la differente prospettiva: non si cerca di conservare per cento o duecento anni, ma si desidera piuttosto risolvere il problema dei prossimi dieci anni, trattando questo periodo non nel suo aspetto contingente, ma interpretandolo come un'istanza all'interno di un processo ciclico.

Un altro aspetto fondamentale del modello è la sua fisicità, riferita non tanto alla sua generica applicabilità anche in ambiente tradizionale – ferma restando un'architettura concettuale dedicata alla conservazione di risorse digitali – quanto all'adozione del bit come unità minima di riferimento. Gli oggetti documentali non sono definiti genericamente come risorse, bensì interpretati e modellati come oggetti concreti, dotati di una loro materialità anche quando siano digitali: l'obiettivo dello standard non è conservare un documento, quanto piuttosto conservare la stringa di bit e le ulteriori informazioni necessarie per rappresentare quel documento. Questo approccio ha sicuramente il merito di scavalcare d'un passo le sottili disquisizioni in merito alla natura e alle caratteristiche essenziali degli oggetti documentali², ma implica la necessità di determinare i confini dei documenti, di

² La comunità archivistica ha sviluppato un ampio filone di ricerca in questa direzione, nel tentativo di riconoscere gli attributi fondamentali dei documenti e – di conseguenza – le modificazioni

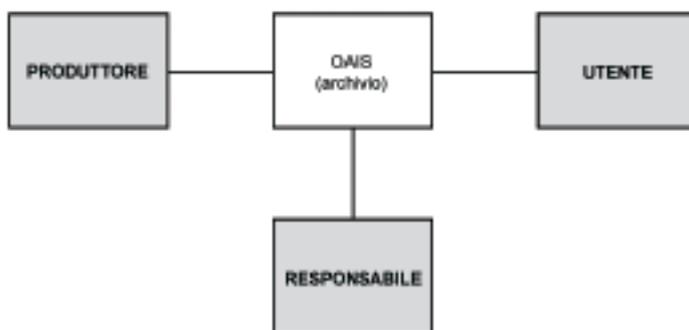
individuare cioè in maniera puntuale le componenti che consentono di rappresentarli nelle forme e nei modi in cui appaiono.

Il modello di riferimento si compone di tre parti:

- la prima descrive l'ambiente all'interno del quale opera un OAIS;
- la seconda ne illustra le componenti funzionali;
- la terza presenta gli oggetti informativi acquisiti, gestiti e messi a disposizione degli utenti.

Ambiente

Un OAIS opera all'interno di un contesto nel quale è possibile individuare i soggetti che a diverso titolo interagiscono con l'archivio:



- Il produttore (*producer*) deposita le risorse informative all'interno di un OAIS affinché siano conservate a lungo termine. Concordate le modalità di versamento nell'archivio mediante un accordo formale (*submission agreement*) che stabilisca il tipo di informazioni e i relativi metadati oggetto di conservazione, nonché i formati da adottare e le modalità operative di trasferimento, il produttore invia il materiale all'OAIS che, attraverso un processo di acquisizione (*ingestion*), accetta le risorse e le predispone per l'inclusione nell'archivio.
- Il responsabile (*management*) si occupa delle politiche complessive di un OAIS

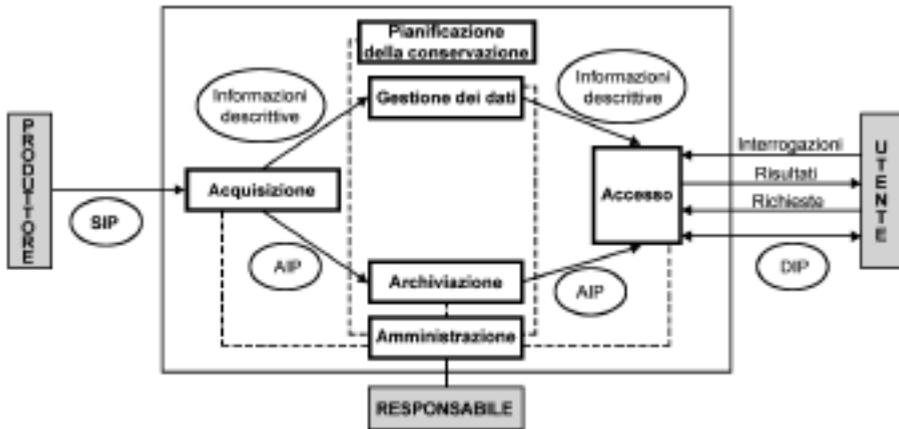
accettabili in sede di conservazione, che non impediscano cioè la verifica dell'autenticità. Tuttavia la letteratura sul tema dell'autenticità dei documenti digitali non è particolarmente ricca, in virtù di un'oggettiva difficoltà a ricondurre entro un quadro teorico stabile l'intrinseca instabilità e varietà degli oggetti digitali. Fra i titoli più significativi merita di essere ricordato quello di Heather MacNeil, *Trusting Records: Legal, Historical and Diplomatic Perspectives*, Dordrecht: Kluwer Academic Publisher, 2000. Il quadro teorico della conservazione e la definizione dei requisiti per l'autenticità hanno avuto uno spazio di riflessione approfondito all'interno del più importante progetto di ricerca promosso in questo ambito: il progetto InterPARES non solo ha proposto il tema dell'autenticità come obiettivo delle indagini di un gruppo internazionale di esperti, ma ha anche istituito una specifica *task force* sull'autenticità, i cui materiali (rapporti, modelli di analisi, studi di casi, questionari, etc.) costituiscono un riferimento fondamentale per gli studi di settore. Per ulteriori informazioni cfr. <http://www.interpares.org>.

- e ne determina l'ambito di sviluppo e le competenze. A tal fine provvede alla pianificazione strategica, alla ricerca dei finanziamenti, alla revisione periodica dei risultati conseguiti e a ogni altra attività gestionale mirata a coordinare lo sviluppo di un OAIS, ma non è coinvolto nelle operazioni quotidiane di amministrazione dell'archivio.
- L'utente (*consumer*) interagisce con l'archivio per cercare e acquisire le informazioni di interesse. La comunità di riferimento (*designated community*) è una particolare classe di utenti caratterizzati dalla capacità di comprendere l'informazione conservata: questo concetto è fondamentale in OAIS, poiché vincola la conservazione all'individuazione dei potenziali utenti del sistema informativo. Di conseguenza è necessario definire e fare riferimento alla base di conoscenza (*knowledge base*) della comunità per determinare cosa conservare: una base di conoscenza che includa la comprensione dell'italiano – ad esempio – non richiederà la conservazione di strumenti di mediazione linguistica come grammatiche e vocabolari qualora la documentazione sia in lingua italiana; diversamente nel caso di una risorsa in giapponese, posto che tale lingua non faccia parte del bagaglio culturale della comunità di base. Il concetto è particolarmente interessante, non solo perché contestualizza la conservazione come un processo rivolto a una comunità di utenti più o meno vasta, di cui si rende necessaria l'identificazione e la caratterizzazione in termini di capacità interpretativa; ma soprattutto perché in tal modo si introduce un fattore dinamico nella funzione conservativa: la base di conoscenza di una comunità non è un dato acquisito una volta per tutte, ma muta continuamente nello spazio e nel tempo e pertanto richiede un monitoraggio continuo onde modificare coerentemente l'oggetto della conservazione. Gli utenti non svolgono quindi un ruolo passivo, costretti a confrontarsi con un archivio alieno dalla propria cultura, ma viceversa è la natura stessa della comunità di riferimento che determina contenuti e modalità di conservazione di un OAIS.

Ovviamente, i ruoli di produttore, responsabile e utente sono funzionali e possono pertanto essere svolti indifferentemente da individui, da enti o da sistemi: non è fuori luogo l'esempio di un OAIS che divenga a sua volta produttore per un altro OAIS, come accade quando la competenza per la conservazione dei materiali passa da un ente all'altro; così come è del tutto normale il caso di un OAIS che svolga il ruolo di utente nei confronti di un altro OAIS. I ruoli presentati hanno cioè natura logica prima ancora che fisica.

Funzioni

Il modello funzionale di un OAIS è basato su sei macro-componenti che svolgono le funzioni fondamentali per garantire la conservazione e l'accesso nel tempo alle risorse informative:



- La funzione di acquisizione (*ingest*) individua l'insieme delle attività finalizzate all'accettazione delle risorse digitali versate dai produttori e alla loro preparazione per l'inserimento nell'archivio. Si articola in un insieme di specifiche sub-funzioni:
 - la presa in carico delle risorse digitali versate dal produttore;
 - il controllo di qualità del materiale ricevuto;
 - la generazione di oggetti digitali che soddisfino i requisiti di conservazione dell'OAIS;
 - la generazione dei metadati descrittivi di supporto alle funzioni di recupero e ricerca dei dati;
 - il trasferimento fisico in archivio degli oggetti digitali predisposti per la conservazione e dei relativi metadati.

La funzione di acquisizione è quindi circoscritta alla fase iniziale del processo di gestione delle risorse, svolgendo un ruolo prevalente di mediazione tra l'OAIS e il produttore, nonché di predisposizione dei dati in maniera adeguata per la conservazione.

- La funzione di archiviazione (*archival storage*) consente la memorizzazione, la tenuta e il recupero delle risorse già elaborate nella fase di acquisizione. L'articolazione delle attività svolte in questo ambito prevede:
 - la presa in carico delle risorse già passate al vaglio dell'acquisizione;
 - la gestione delle priorità di archiviazione, grazie alla quale si individua l'adeguato supporto di memorizzazione sulla base di politiche generali, osservazioni statistiche o precise direttive;
 - la sostituzione periodica dei supporti;
 - il controllo degli errori, per verificare in maniera statisticamente accettabile che le risorse non siano corrotte nelle fasi di trasferimento interno;
 - la prevenzione dei danni in caso di disastro, mediante la duplicazione di tutti gli oggetti digitali presenti nel sistema e la loro archiviazione in altro luogo;

- la fornitura di una copia delle risorse archiviate, in caso di richiesta di accesso.
- La funzione di gestione dei dati (*data management*) è in un certo senso il cuore dell'OAIS, poiché controlla il database contenente i metadati descrittivi che identificano e descrivono le risorse archiviate, nonché i dati relativi alle operazioni di amministrazione interna (come le statistiche di accesso o il controllo delle prestazioni). Le specifiche e limitate sub-funzioni di amministrazione, interrogazione, generazione di rapporti e ricezione degli aggiornamenti sono quindi tutte inerenti complessivamente alla gestione della base di dati.
- La pianificazione della conservazione (*preservation planning*) è una funzione che ha l'obiettivo di progettare le strategie di conservazione dell'OAIS e di aggiornarle in relazione ai cambiamenti dell'utenza o delle tecnologie. Richiede pertanto un costante intervento per:
 - il controllo sistematico della comunità di riferimento, al fine di individuare eventuali nuove esigenze (formati, supporti, software, piattaforme operative) mediante censimenti, processi di verifica formale, workshop e qualunque altra iniziativa che favorisca una risposta dell'utenza;
 - il controllo sistematico delle tecnologie, con l'obiettivo di individuare quelle emergenti e quelle soggette a fenomeni di obsolescenza;
 - lo sviluppo di strategie e standard per la conservazione;
 - lo sviluppo di piani di migrazione e di nuovi requisiti per la conservazione nell'OAIS.
- La funzione di accesso (*access*) consente agli utenti di acquisire informazioni sulle risorse conservate in archivio ed eventualmente di richiederne copia. Le azioni svolte in questo contesto sono articolate in tre aree:
 - il coordinamento delle attività di accesso, che fornisce all'utente un'interfaccia unitaria per accedere al materiale posseduto dall'archivio, grazie a semplici interrogazioni (*query*) che producono una lista di risultati da analizzare o a vere o proprie richieste (*order*) di copia della fonte di interesse;
 - la generazione della copia, che recupera la risorsa archiviata e ne realizza una copia associata ai relativi metadati descrittivi, effettuando tutte le operazioni necessarie (conversioni, elaborazioni di immagini, ecc.) per consentirne la corretta rappresentazione;
 - la consegna della risposta, che agisce da filtro prima dell'inoltro all'utente dei risultati delle interrogazioni o delle richieste.

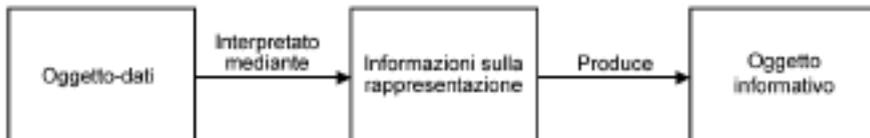
La funzione di accesso è quindi circoscritta alla fase finale del processo di gestione delle risorse, svolgendo un ruolo prevalente di mediazione tra l'OAIS e l'utente, nonché di predisposizione dei dati in maniera adeguata per la diffusione.

- La funzione di amministrazione (*administration*) cura la gestione delle operazioni quotidiane di un OAIS al fine di garantirne l'operatività. Afferiscono a questo ambito la definizione delle modalità di versamento delle risorse nell'archivio, la gestione della configurazione complessiva del sistema, l'aggiornamento delle informazioni e il controllo dell'accesso, nonché il servizio per i clienti, finalizzato a soddisfare i bisogni informativi di carattere generale e a gestire le informazioni relative a pagamenti e fatturazioni.

Non è forse inutile evidenziare come il modello OAIS risponda ovviamente a logiche formali di descrizione che, al di là dei tecnicismi e delle inevitabili approssimazioni spesso dovute alla necessità di inscrivere la complessità del reale all'interno di uno schema discreto, mirano tuttavia a tradurre prassi e concetti talora già consolidati nelle discipline del documento: non è difficile riconoscere in alcune delle funzioni illustrate le attività che compongono la quotidianità della vita professionale di un bibliotecario o di un archivista.

Oggetti informativi

Il modello OAIS è decisamente orientato verso un approccio atomistico: l'oggetto informativo (*information object*) è infatti frammentato e ridotto ai minimi termini, distinguendo in esso i dati dalle informazioni sui dati, interpretandolo cioè come composizione di un oggetto-dati (*data object*) e delle relative informazioni sulla rappresentazione (*representation information*) che ne permettono la comprensione.



Questo approccio ha conseguenze di enorme rilievo sul profilo conservativo, poiché l'inserimento di una risorsa all'interno di un OAIS implica l'identificazione e la comprensione delle due componenti. In ambito digitale ciò significa distinguere i bit che costituiscono l'oggetto-dati dalle relative informazioni di rappresentazione: ad esempio, un semplice file ASCII contenente un elenco di numeri dovrà assumere tra le informazioni di rappresentazione la tabella di codifica ASCII che consenta di rintracciare le corrispondenze tra bit e simboli alfanumerici; ma sarà anche necessario indicare che i numeri sono – supponiamo – delle coppie di valori e vanno pertanto letti in associazione; che si riferiscono – ad esempio – ad altezza e peso di una popolazione di soggetti; e che sono espressi in determinate unità di misura. E così via.

Il modello muove quindi dai dati per costruire su di essi una struttura concentrica di metadati tramite il meccanismo ricorsivo: i dati richiedono altri dati per poter

essere correttamente interpretati; questi a loro volta possono richiedere altri dati interpretativi; e così via, finché non si raggiunge un livello autosufficiente, determinato dalla base di conoscenza della comunità di riferimento. Qui si arresta un'iterazione che altrimenti condurrebbe paradossalmente alla conoscenza del mondo: ad esempio, nel caso di un testo italiano in formato digitale, trascurando per brevità tutti gli altri aspetti, non sembra necessario annoverare un vocabolario tra le informazioni di rappresentazione; ma se ci si pone nell'ottica della conservazione a lungo termine un orizzonte temporale così ampio può giustificare la conservazione delle informazioni relative a semantica e sintassi del linguaggio naturale, data la possibilità di evoluzioni significative nell'uso. Il problema è determinare il momento in cui la cultura dell'utenza di riferimento non è più adeguata a interpretare un oggetto con le consuete informazioni di rappresentazione: quando è necessario introdurre questo ipotetico vocabolario? fra cinquanta anni? o cento? La risposta è ovviamente vincolata alle caratteristiche della comunità di riferimento, ma l'aspetto che qui si vuole sottolineare è la dinamicità della funzione conservativa, che richiede una sensibilità costante nei confronti dell'utenza al fine di aggiornare la risorsa archiviata.

E vale forse la pena sottolineare una caratteristica che emerge chiaramente da questa impostazione metodologica: il processo conservativo deve essere applicato tanto ai dati quanto ai metadati, poiché gli uni e gli altri sono interpretati come oggetti informativi, caratterizzati da un contenuto informativo e da informazioni sulla rappresentazione. Anzi, paradossalmente occorre prestare maggiore attenzione al complesso sistema di metadati "agganciato" al dato, poiché è solo da questi che emerge il significato del dato, altrimenti inerte. In questo senso, l'intero modello OAIS può essere analizzato e interpretato allo scopo di individuare i metadati soggiacenti e costruire un sistema di elementi informativi funzionali alla conservazione a lungo termine³.

A partire dalla fondamentale categorizzazione che disgiunge l'oggetto-dati dalle informazioni sulla rappresentazione, il modello OAIS definisce il pacchetto di informazioni (*package information*), l'entità fondamentale attorno alla quale ruotano i processi di conservazione, un insieme logico composto dall'oggetto digitale che deve essere conservato e dai metadati necessari a garantirne la conservazione e l'accesso a lungo termine. In pratica, l'oggetto informativo obiettivo della conservazione assume il ruolo di contenuto e il pacchetto informativo risulta così composto da due componenti: le informazioni sul contenuto (*content information*) e le informazioni sulla conservazione (*preservation description infor-*

³ Il gruppo di ricerca internazionale Preservation Metadata: Implementation Strategies (PREMIS) nasce nel 2003 con l'obiettivo di definire dettagliatamente un set di metadati per la conservazione utilizzabile in generale nell'ambito della *digital preservation*, a partire dal modello OAIS. I risultati sono illustrati in *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group*, May 2005, OCLC-RLG, 2005. Cfr. <http://www.oclc.org/research/projects/pmwg>.

mation). Le informazioni di impacchettamento (*packaging information*) si riferiscono invece all'identificazione e alla descrizione del pacchetto informativo: è il legante delle varie componenti, che permette loro di essere identificate e localizzate come una singola unità logica.



Le informazioni descrittive (*descriptive information*) permettono all'utente di recuperare il contenuto di interesse inserito all'interno di un pacchetto informativo: la loro complessità varia dal semplice titolo all'insieme strutturato di attributi (come nel caso di un record Dublin Core).

Esistono tre diverse tipologie di pacchetto informativo:

- il SIP (*submission information package*), per il versamento;
- l'AIP (*archival information package*), per l'archiviazione;
- il DIP (*dissemination information package*), per la distribuzione.

Il pacchetto di versamento è il pacchetto informativo inviato dal produttore e gestito in fase di acquisizione. Il suo formato e i suoi contenuti sono normalmente concordati tra l'OAIS e il produttore: il controllo di qualità effettuato all'ingresso è finalizzato appunto a verificare il rispetto dei requisiti stabiliti, la cui incompletezza o inadeguatezza può essere risolta con l'integrazione a cura dell'OAIS o con la richiesta al produttore di un nuovo versamento. Spesso sono necessari più pacchetti di versamento per comporre in maniera unitaria il contenuto e le informazioni sulla conservazione di un pacchetto di archiviazione; viceversa, è possibile che le informazioni di un singolo pacchetto di versamento debbano essere incluse in più pacchetti di archiviazione.

All'interno dell'OAIS, i pacchetti di versamento sono trasformati in pacchetti di archiviazione nella fase di acquisizione: la risorsa archiviata e i relativi metadati di conservazione costituiscono un singolo pacchetto logico atto a garantire la ge-

stione e la permanenza dei dati nel sistema. Il pacchetto di archiviazione può consistere di una singola unità (*archival information unit*) o essere in realtà un'aggregazione di più pacchetti archivistici (*archival information collection*): ciò rende possibile la gestione uniforme di risorse accomunate da caratteristiche logiche o fisiche simili.

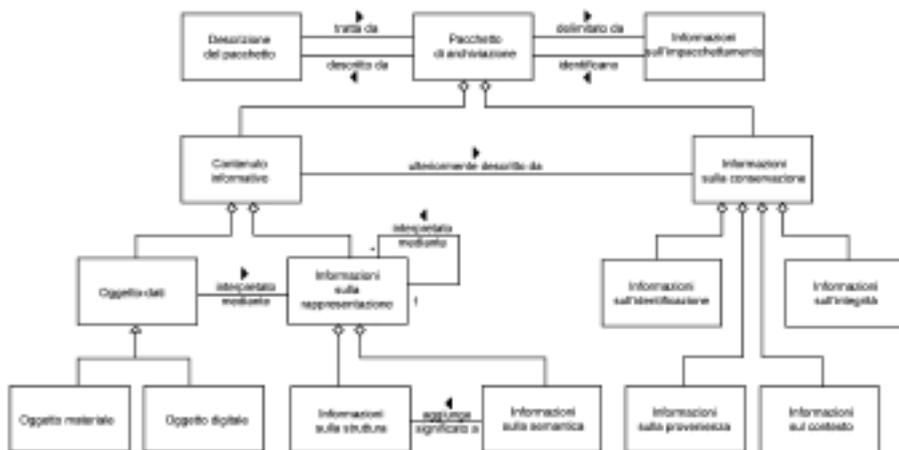
Il pacchetto di distribuzione è il pacchetto informativo consegnato all'utente in risposta alle sue richieste di accesso. Può riferirsi a un singolo pacchetto di archiviazione oppure a una raccolta di pacchetti archivistici, e le informazioni sulla conservazione fornite all'utente possono essere complete o meno, sulla base delle politiche stabilite dall'OAIS.

L'aspetto più rilevante che emerge da questo modello è la chiara affermazione della necessità di una differenziazione tra oggetti acquisiti, archiviati e presentati all'utente: le diverse esigenze collegate alle differenti fasi di gestione delle risorse nell'OAIS determinano trasformazioni della struttura, del formato e dei contenuti degli oggetti digitali (un esempio banale è la conversione di un file TIF in JPG nel passaggio da pacchetto di archiviazione a pacchetto di distribuzione, in maniera tale da ridurre i tempi di trasferimento dei dati all'utente).

Le informazioni sulla conservazione associate al contenuto (*content information*) sono articolate in quattro aree riferite ad altrettanti profili di rilievo in sede conservativa:

- le informazioni sull'identificazione (*reference information*) forniscono un identificatore univoco per il contenuto, come ad esempio un codice ISBN o un URL;
- le informazioni sul contesto (*context information*) descrivono le relazioni tra il contenuto del pacchetto informativo e i contenuti di altri pacchetti;
- le informazioni sulla provenienza (*provenance information*) documentano la storia del contenuto, cioè la sua creazione, le trasformazioni subite, i passaggi di possesso o di proprietà;
- le informazioni sull'integrità (*fixity information*) consentono di proteggere il contenuto da alterazioni non autorizzate, ad esempio con l'utilizzo di *checksum* o di strategie più sofisticate come impronte e firme digitali.

L'architettura che risulta da questo quadro è sicuramente complessa, ma estremamente efficace per l'analisi degli elementi minimi che partecipano ai processi conservativi. La rigorosa frammentazione che sostiene l'intera struttura del pacchetto informativo consente ovviamente un dettaglio esasperato degli oggetti coinvolti nei processi di gestione documentaria, ma permette al contempo di descrivere i vincoli che identificano i *corpora* documentari: è evidente che per "salire" dal bit al fascicolo e alla serie, o al libro e alla collezione, è necessario adottare un meccanismo ricorsivo e praticarlo numerose volte, considerando iterativamente un pac-



chetto come oggetto informativo di un pacchetto di livello superiore. Un approccio certo macchinoso e complicato da gestire, ma con il vantaggio di un rigore che tutela la funzione conservativa definendone gli aspetti fisici e logici.

Progetti di ricerca

Nel corso degli ultimi anni vari e importanti progetti di ricerca nell'ambito della conservazione digitale hanno assunto OAIS come modello di riferimento. La prima e più famosa esperienza è sicuramente quella di NEDLIB (Networked European Deposit Library), un progetto di collaborazione tra diverse biblioteche europee finalizzato alla realizzazione di un'infrastruttura sulla quale costruire un deposito librario condiviso per la conservazione permanente delle pubblicazioni in formato digitale⁴. Realizzata tra il 1998 e il 2000, la ricerca ha condotto all'elaborazione di un vero e proprio sistema di deposito per le pubblicazioni elettroniche (DSEP – Deposit System for Electronic Publications) fondato sui concetti e sulle riflessioni presenti nel modello OAIS: l'architettura delle funzioni, le strutture di metadati, perfino le valutazioni dei software realizzati nel corso della ricerca assumono lo standard OAIS come riferimento, con l'auspicio di una progressiva diffusione e di un accoglimento favorevole da parte del mercato, a vantaggio dello sviluppo di sistemi aperti e di prodotti di qualità⁵. Analoghe le finalità del progetto DSPACE⁶ avviato nel 2000 dal Massachusetts Institute of Technology per la realizzazione di un deposito digitale coerente con i principi e i modelli definiti in OAIS: il software sviluppato nell'ambito di questo progetto è liberamente disponibile⁷ dal 2002 per le

⁴ Cfr. <http://nedlib.kb.nl>.

⁵ In verità occorre rilevare che il progetto NEDLIB si discosta significativamente dal modello OAIS, proponendo una strategia conservativa fondata sull'emulazione piuttosto che sulla migrazione, e integrando il modello funzionale con un affinamento dei processi di acquisizione, di distribuzione e di conservazione fisica.

⁶ Cfr. <http://www.dspace.org>.

⁷ Sotto licenza Berkeley OpenSource Software Distribution.

istituzioni che vogliono creare un deposito digitale, con l'obiettivo di favorire la realizzazione di un deposito federato di istituti di ricerca. Identica la funzione del DRS (Digital Repository Service), un sistema sviluppato a partire dal 1998 per il deposito di materiali digitali presso l'Università di Harvard⁸. Negli stessi anni – con una dilatazione fino al 2002 – il progetto CEDARS (CURL Exemplars in Digital Archives) si concentra sui problemi della conservazione digitale con l'obiettivo di elaborare documenti e prodotti rivolti non solo alle istituzioni aderenti al CURL (Consortium of University Research Libraries), ma anche agli operatori del settore⁹. Il set di metadati per la conservazione, presentato in un'autonoma guida, fa esplicito riferimento alle informazioni sulla conservazione così come definite e articolate in OAIS; il prototipo di archiviazione digitale distribuita è dotato di caratteristiche peculiari, ma ha un deciso radicamento nelle strutture OAIS; e anche la guida alle strategie di conservazione digitale contiene richiami ai concetti OAIS. Analoga l'influenza dello standard su altre iniziative, tra le quali segnaliamo CAMILEON (Creative Archiving at Michigan & Leeds: Emulating the Old on the New), in stretto collegamento con CEDARS, ma dedicata a un'analisi specifica dell'emulazione come strategia primaria di conservazione¹⁰; PANDORA (Preserving and Accessing Networked Documentary Resources of Australia), l'archivio Web australiano che raccoglie le pubblicazioni australiane on-line, dotato di un sistema di acquisizione e gestione distribuita delle risorse digitali¹¹; ERA (Electronic Records Archive), il progetto proposto dall'amministrazione archivistica statunitense per la conservazione dei documenti digitali¹²; ARSBNi (Arricchimento dei servizi della bibliografia nazionale italiana), un progetto di digitalizzazione di frontespizi e indici della produzione libraria italiana promosso dalla Biblioteca nazionale di Firenze¹³; e BIBIT (Biblioteca italiana), una raccolta digitale di testi rappresentativi della tradizione culturale e letteraria italiana dal Medioevo al Novecento¹⁴. Uno spazio a parte merita il progetto CASPAR (Cultural, Artistic and Scientific knowledge for Preservation, Access and Retrieval)¹⁵ in ragione dell'ampiezza degli obiettivi che si propone. Avviato nel 2006 nell'ambito del sesto programma quadro dell'Unione Europea e destinato a concludersi nel 2009, il progetto mira a costruire un'infrastruttura per la conservazione mediante:

⁸ Cfr. <http://hul.harvard.edu/ois/systems/drs>.

⁹ Cfr. <http://www.leeds.ac.uk/cedars>.

¹⁰ Cfr. <http://www.si.umich.edu/CAMILEON>.

¹¹ Cfr. <http://pandora.nla.gov.au>.

¹² Cfr. <http://www.archives.gov/era>.

¹³ Cfr. <http://www.bncf.firenze.sbn.it/progetti/arsbni/>.

¹⁴ Cfr. <http://www.bibliotecaitaliana.it/>.

¹⁵ Cfr. <http://www.casparpreserves.eu>. L'Autore partecipa – con l'Università di Urbino – al progetto CASPAR con lo specifico obiettivo di indagare il profilo dell'autenticità degli oggetti digitali al fine di promuovere lo sviluppo di strategie e strumenti atti a garantire una conservazione compatibile con le esigenze di autenticità. Il gruppo di ricerca dell'Università di Urbino è diretto da Mariella Guercio ed è composto da Cecilia Castellani, Monica Grossi, Giovanni Michetti.

- l'individuazione di una metodologia applicabile a dati e applicazioni afferenti al contesto scientifico, artistico e più latamente culturale;
- la ricerca e lo sviluppo di componenti tecnologiche integrate da utilizzare come base per la realizzazione dell'infrastruttura, intesa come piattaforma software per la costruzione di servizi e applicazioni che possano essere adattati a specifiche aree di interesse.

A tal fine, CASPAR assume esplicitamente OAIS come modello su cui fondare l'architettura dell'intera infrastruttura e vincola la costruzione delle singole componenti software al rispetto dell'impostazione concettuale fornita dallo standard. La realizzazione di specifiche sperimentazioni nei diversi ambiti oggetto di ricerca è funzionale sia all'individuazione delle criticità specifiche e comuni che affliggono la conservazione delle risorse digitali, sia alla verifica delle metodologie e delle soluzioni tecnologiche proposte, al fine di garantirne la replicabilità in ambiti diversi. Ulteriore elemento di interesse è la scelta di spostare l'attenzione dai dati alla conoscenza, con l'obiettivo di assicurare la conservazione nel futuro degli archivi nonché dei sistemi e dei servizi correlati: a tal fine il progetto enfatizza il ruolo delle ontologie espresse mediante linguaggi formali come mezzo per lo scambio e la conservazione dei sistemi di conoscenze.

Conclusioni

Il modello OAIS, a dispetto delle sue evidenti caratteristiche di astrazione, è stato proposto come riferimento all'interno di vari e importanti progetti applicativi. Occorre tuttavia riconoscere che, allo stato attuale, non esistono implementazioni complete del modello, utilizzato più come stella polare, come *framework* entro il quale ricondurre idee e oggetti, che come *checklist* per la realizzazione di un sistema applicativo integrato. Esistono certo specifiche soluzioni e moduli applicativi che possono dichiararsi a buon diritto *OAIS compliant*, per lo più in relazione a determinate funzioni, ma la costruzione di un vero e proprio archivio OAIS è impresa che richiede ancora investimenti scientifici ed economici di rilievo. Al contempo, è necessario armonizzare gli sforzi condotti su più fronti ed estendere il modello per costruire un ambiente ove strumenti, oggetti e concetti rispondano a una logica unitaria: significativa in questo senso l'iniziativa del CCSDS¹⁶ che ha avviato i lavori per l'elaborazione di uno standard per la certificazione dei *digital repository*, basandosi su fonti già disponibili e coerenti con il modello OAIS¹⁷, ma non dimenticando né il generico riferimento agli standard ISO 9000 e ISO 27000 né le ovvie istanze di compatibilità con gli altri documenti elaborati a supporto del modello OAIS¹⁸.

¹⁶ Più precisamente, del Digital Repository Audit and Certification Working Group.

¹⁷ Si vedano ad esempio i risultati del progetto NESTOR all'indirizzo <http://www.langzeitarchivierung.de/index.php>.

¹⁸ Si veda ad esempio il *Producer-Archive Interface Methodology Abstract Standard*, Blue Book, May 2004, disponibile sul sito del CCSDS, <http://public.ccsds.org/publications/archive/651x0b1.pdf>

Come già detto, i problemi della conservazione digitale sono numerosi e complessi: lungi dal rappresentare la risposta definitiva, il modello OAIS è tuttavia non solo una proposta metodologica, ma soprattutto una fonte di stimolo ormai solida per la costruzione di una cultura della conservazione digitale con cui governare la sedimentazione di oggetti e processi in ambiente digitale.

Recent years have seen the development of numerous initiatives devoted to identifying solutions for long-term digital preservation. Some of those initiatives have been aimed at defining metadata element sets to manage information elements needed to support the preservation process; some have been studying techniques for transferring information from an obsolete context to new media, systems and environments; some have analysed the costs and benefits of different preservation strategies; while others have engaged with theoretical analysis, striving to identify the crucial elements of digital records. A complex and urgent matter such as digital preservation indeed requires simultaneous action on all fronts, spanning from the technical to the theoretical side, and dealing with both organizational and socio-cultural aspects: to build a preservation policy we need to promote awareness and active participation amongst a diversified range of stakeholders, outreaching well beyond the narrow borders of the domain community.

The Open Archival Information System (OAIS) reference model is a conceptual framework for describing objects, processes, strategies, and techniques for long-term digital preservation; it also helps understanding their relationships, and analysing and comparing different preservations solutions. The model is based on a dynamical approach to preservation, seen as a permanent, never-ending process aimed at monitoring the context in which objects are plunged into; and it designs a complex architecture of functions and information objects, founded on the identification of the resources needed to recreate the meaning of objects, assuming the bit as the minimal unit of the conceptual system.

Durant les dernières années, les initiatives visant à repérer des solutions pour la conservation numérique à long terme se sont multipliées. Parmi celles-ci, certaines cherchent à définir les structures des métadonnées afin de gérer les éléments informatifs nécessaires au processus de conservation, d'autres sont destinées à l'étude des techniques de transfert de l'information, depuis un contexte obsolète vers de nouveaux supports coût-bénéfices reliés aux différentes stratégies de conservation; des initiatives, toutes aussi importantes, ont approfondi l'analyse théorique afin de repérer les composantes significatives des objets documentaires. En réalité, le problème de la conservation numérique est si complexe et pressant qu'il faut agir dans toutes les directions à la fois, en allant de la dimension technique à la dimension théorique, sans délaisser le profil de l'organisation et le profil socioculturel. En effet, la construction d'une politique de conservation passe inévitablement à travers la sensibilisation et l'engagement des sujets de la société qui ne sont pas renfermés dans les confins étroits de la communauté de domaine.

Le modèle de référence OAIS (Open Archival Information System) se propose comme un cadre conceptuel unitaire visant à décrire les objets, stratégies et techniques finalisés à la conservation numérique à long terme, ainsi qu'à comprendre leurs relations réciproques et à analyser et confronter différentes solutions de conservation. Le modèle est fondé sur une idée dynamique de la conservation, conçue comme un processus permanent et infini de monitoring du contexte dans lequel les objets se trouvent; ce modèle dessine une architecture de fonctions et d'objets informatifs complexe basée sur l'individuation des ressources nécessaires à la reconstruction de la signification des objets, et il a adopté le bit comme l'unité la plus petite du système conceptuel.