

Gloria Ciocchi
ciocchi_g@camera.it

http://ssdoo.gsfc.nasa.gov/nost/isoas/ref_model.html

Se, tradizionalmente intesa, la *conservazione* consiste nell'assicurare la stabilità nel tempo dei supporti fisici, in modo da assicurare l'accesso all'informazione da essi veicolata, oggi la crescente quantità di informazione in formato elettronico ha spostato l'attenzione su nuovi fattori di rischio, legati non solo alla labilità dei supporti, ma al rischio di veloce obsolescenza di sistemi operativi e formati.

Le caratteristiche stesse dell'informazione digitale, poi, estendono le preoccupazioni riguardo alla conservazione e alla durata dal tradizionale ambito delle biblioteche e degli archivi a settori finora immuni da tali preoccupazioni, quali organizzazioni del settore governativo, commerciale e non-profit.

Molta dell'informazione necessaria a conservare questi dati è disponibile più facilmente, o addirittura soltanto, nel momento stesso della loro creazione: ne consegue che il controllo istituzionale sui meccanismi di archiviazione si sposta a monte, ed investe il produttore stesso dell'informazione.

Per questo motivo, nell'ambito dei lavori del sottogruppo **MAG**, ci è sembrato utile, per individuare i **requisiti funzionali di un set di metadati gestionali-amministrativi e strutturali che abbia applicabilità generale**, partire dalla definizione delle componenti chiave e dei processi su cui si basano le attività di archiviazione, gestione e conservazione di documenti elettronici, sulla scia di alcune importanti esperienze, nazionali e internazionali, nel campo dei metadati, che hanno adottato il **modello di tipo logico-funzionale denominato OAIS** (Open Archival Information System).

Riferirsi ad un modello comune, condividere una terminologia e una cornice concettuale facilita lo scambio di idee e la comparazione di esperienze, non solo tra biblioteche, ma anche tra istituzioni diverse, quali musei, archivi, enti governativi: la funzione svolta dal modello OAIS nel contesto dei metadati gestionali-amministrativi e strutturali può essere paragonata alla funzione svolta da FRBR per i metadati relativi alla ricerca e al recupero delle risorse informative.

L'acronimo OAIS e' traducibile con **Sistema Informativo Aperto per l'Archiviazione**: si tratta di uno standard ISO in fase di elaborazione attraverso un processo, per l'appunto, "aperto" ai contributi di enti e istituzioni diverse. L'ISO ha incaricato del coordinamento dell'iniziativa il **Consultative Committee for Space Data Systems** della NASA, in relazione allo sviluppo di standards atti alla conservazione a lungo termine dei dati derivati dall'osservazione dello spazio e del pianeta.

Benché sviluppato in origine dalla comunità della ricerca spaziale, il modello OAIS è adeguato alle finalità della conservazione a lungo termine anche per altri tipi di comunità: infatti, pur essendo fortemente orientato al trattamento dei documenti elettronici, è applicabile a qualsiasi tipo di archivio - digitale o analogico; esso non specifica nessun tipo di implementazione e può essere applicato indifferentemente a oggetti digitali nativi, a prodotti di attività di digitalizzazione (quali files di immagini), e persino a oggetti fisici.

Di fatto, l'OAIS si sta rapidamente affermando come lo standard emergente per la conservazione delle risorse digitali.

Ciò che esso accentua fortemente, infatti, è la tematica della *responsabilità* della conservazione; in modo particolare, della conservazione a lungo termine dell'informazione digitale: ed è per questi motivi che è stato adottato come modello di riferimento da alcune importanti esperienze del mondo di biblioteche ed archivi.

Tra di esse può essere citata in primo luogo NEDLIB - Networked European Deposit Library, che ha formulato un modello di deposito delle pubblicazioni elettroniche, poi il progetto CEDARS (CURL Exemplars in Digital Archives), promosso dalle Università di Cambridge, Oxford e Leeds, che ha sviluppato uno schema di metadati per la conservazione basato sul modello OAIS, e PANDORA (Preserving and Accessing Networked Documentary Resources of Australia) della National Library of Australia, che mira alla conservazione di siti Web selezionati. Ad esso si è ispirata implicitamente la Library of Congress per la definizione del proprio set di metadati, e la BNF per il progetto ARSBNI.

L'OAIS si autodefinisce *"un'organizzazione di soggetti e sistemi che hanno accettato la responsabilità della conservazione dell'informazione e del mantenerla disponibile per una comunità determinata"*; come tale, si propone di fornire una solida base per una ulteriore standardizzazione nel contesto dell'archiviazione digitale e di promuovere una maggiore consapevolezza, da parte dei venditori di tecnologie, delle esigenze della conservazione a lungo termine dei documenti elettronici.

Riconoscendo che le collezioni digitali sono caratterizzate da una natura fortemente distribuita, e che vi è necessità, al tempo stesso, di sviluppare localmente politiche e procedure efficaci di gestione e conservazione di tali risorse, esso delinea un modello di archiviazione distribuita, ma rispondente ad un ben individuato modello logico.

A tale scopo esso individua termini e concetti rilevanti per l'archiviazione di documenti digitali, identifica le componenti e i processi chiave comuni alla maggior parte delle attività di conservazione digitale, e propone un modello **logico di riferimento** per gli oggetti digitali e i metadati loro associati, che comprende la creazione e l'uso dei metadati utili a gestire il materiale elettronico, dalla fase di acquisizione a quella della conservazione, fino alla fase di accesso.

Un'altra caratteristica del modello è che non esprime alcuna preferenza riguardo alle strategie di conservazione dei dati da adottare (p. es. migrazione o emulazione).

Vorrei illustrare sinteticamente i principali concetti su cui il modello si basa, e che sono stati ripresi dai progetti menzionati:

AMBIENTE E COMPONENTI FUNZIONALI

(LUCIDO 1)

L'**ambiente** OAIS è formato dall'interazione di quattro **entità**, definite come:

1. **i produttori** (che sulla base di accordi formalizzati per l'immissione dei dati nell'archivio forniscono i dati in base alle componenti logiche e al modello di rappresentazione OAIS)
2. **i consumatori** (le persone, o sistemi clienti, che interagiscono con i servizi OAIS per reperire e acquisire l'informazione conservata di proprio interesse. Tra i consumatori la

Comunità Designata costituisce un tipo particolare di Consumatore, poiché si presuppone comprenda l'informazione nella forma in cui viene archiviata).

3. **il management** (esterno all'archivio, si occupa delle politiche dell'archivio: cosa archiviare, come trovare i fondi ecc.)
4. **l'archivio stesso** (inteso come organizzazione che si propone di conservare l'informazione per consentirne l'accesso e l'uso ad una Comunità Designata)

Le **componenti funzionali** dell'OAIS sono definite come:

1. **Immissione** (Ingest) in questa fase si riceve l'informazione dai produttori e la si prepara per l'archiviazione
2. **Archiviazione dei dati** (Archival Storage): tratta l'archiviazione, la manutenzione e la gestione della informazione archiviata
3. **Gestione dei dati** (Data Management): coordina i metadati *descrittivi* relativi sia all'informazione archiviata *che ai dati amministrativi interni all'archivio*
4. **Accesso** (Access): è la funzione che aiuta i consumatori a identificare e ottenere informazione dall'archivio
5. **Amministrazione dell'archivio** (Administration): ha in carico le operazioni giornaliere di mantenimento dell'archivio.

A queste cinque componenti funzionali si aggiungerà, nella revisione del draft che è in corso, una sesta componente

6. **Pianificazione della conservazione** (Preservation Planning)

proposta dal progetto NEDLIB: esso, infatti, rilevando come il modello OAIS, pur esprimendo implicitamente una preferenza verso la migrazione dei dati come strategia di conservazione, non specificasse poi concretamente quali procedimenti fossero necessari e quale entità funzionale vi venisse coinvolta, (se si prescinde dalle operazioni di *refreshing* e di *copying* su nuovi media attribuite alla funzione di Archiviazione), ha proposto una sesta entità funzionale, **Pianificazione della Conservazione**, che specifica e rende "visibili" ulteriormente le attività di conservazione effettuate, a prescindere dalla strategia adottata.

CLASSIFICAZIONE DEGLI OGGETTI INFORMATIVI

Oltre alle componenti funzionali, l'OAIS fornisce anche un **modello di strutturazione dei dati** adeguato a **rappresentare l'informazione** digitale da un punto di vista orientato alla conservazione. Centrale è la definizione di **Oggetto Informativo** (Information Object), di qualsiasi tipo esso sia, come composto da due elementi:

- **Dati**
- **Informazione sulla Rappresentazione**, necessaria per conferire significato ed interpretabilità ai dati.

La conservazione nel tempo dei due elementi è un requisito fondamentale nel funzionamento del modello.

In un ambiente digitale, i **Dati** (una sequenza di bits) sono necessariamente collegati ad una **R.I.** (Representation Information/ **Informazione sulla Rappresentazione**) che contiene tutto ciò che è necessario per rendere comprensibili quei bits, sia dal punto di vista strutturale

(specificazione del formato, descrizione del s/w di accesso, etc.), sia semantico (p. es. in quale lingua è un testo in caratteri ASCII).

Ma anche per un documento fisico vale lo stesso modello di strutturazione; per es., l'informazione contenuta in un libro in italiano, espressa da caratteri a stampa (i **dati**), in combinazione con una conoscenza della lingua del testo (la Conoscenza di Base) si converte in informazione. Se la lingua italiana non rientra tra le conoscenze del ricevente, il testo (i dati) dovrà essere accompagnato da un dizionario italiano e da informazioni grammaticali espresse in una forma compresa nella Conoscenza di Base di chi riceve l'informazione: questi costituiranno **l'Informazione sulla Rappresentazione**.

In generale, perciò si può dire "*i Dati, in unione alla loro Informazione sulla Rappresentazione, veicolano significato*".

E' da notare, tuttavia, che la Informazione sulla Rappresentazione, in ambiente digitale, è dotata di una natura ricorrente: per cui una I.R., per essere interpretata, avrà bisogno di un'altra I.R., e così via, creando una Rete di Informazioni sulla Rappresentazione, (Representation Network), fino ad incontrare la Conoscenza di Base della Comunità per cui l'informazione viene conservata. Il modello OAIS raccomanda che la Rete termini con un documento in formato analogico che dia avvio all'intera sequenza di interpretazione .

LUCIDO 2 (Pacchetti di Informazione)

Ogni scambio di informazione, da e per l'archivio, e all'interno di OAIS, avviene attraverso l'utilizzazione di **Pacchetti di Informazione** (I.P.-Information Packages), che sono dei contenitori concettuali di dati.

Gli **Oggetti Informativi** possono comporre 3 tipi di **I.P.**, utilizzati a seconda che il flusso di informazioni avvenga dal Produttore all'OAIS, al suo interno, o dall'archivio stesso verso il Consumatore:

- **SIP** - (*Submission Information Package*) - **Pacchetto di Informazioni per l'Immissione**, utilizzato nella fase di immissione/acquisizione dei dati, mandato dal produttore in base al Submission Agreement stipulato con l'OAIS
- **AIP** - (*Archival Information Package*) - **Pacchetto di Informazioni per l'Archiviazione**, destinato alla conservazione a lungo termine.
- **DIP** - (*Dissemination Information Package*) - **Pacchetto di Informazioni per la Distribuzione**, trasferito dall'OAIS all'utente in base ad una richiesta di accesso.

Un **I.P.** è costituito dall'aggregazione di quattro classi di **Oggetto Informativo**:

1. **Informazione sul Contenuto** / Content Information: l'oggetto primario destinato ad essere conservato dall'archivio (p.es. un documento in formato PDF (i dati), più la documentazione del formato PDF (la R.I.)
2. **Informazione Descrittiva per la Conservazione** / Preservation Description Information. E' il set di informazioni necessarie a conservare adeguatamente, per un periodo di tempo indefinito, il Contenuto cui sono associate; si focalizza sulla descrizione degli stati passati e presenti del Contenuto, assicura che sia identificato univocamente, e che non sia alterato senza che ciò venga registrato.

Il Contenuto e l'Informazione Descrittiva per la Conservazione (PDI) sono viste come incapsulate e identificabili per mezzo della:

3. **Informazione sulla Composizione del Pacchetto di Informazione** ovvero **Informazione di Impacchettamento** (Packaging Information): come reperire **1) e 2)**: ovvero come sono collegati i componenti di un Pacchetto di Informazioni in un'entità identificabile su uno specifico supporto (p. es. in che disco, in che directory ecc.)

Il “pacchetto di informazione” che ne risulta è ricercabile attraverso la

4. **Informazione Descrittiva** (Descriptive Information): finalizzata alla ricerca e al recupero dell'informazione, p. es.: Manzoni, Alessandro - Autore principale - RICA. Essa può basarsi su informazioni contenute nell'Informazione sul Contenuto e nell'Informazione Descrittiva per la Conservazione, ma ne è distinta sul piano logico.

LUCIDO 3 (PDI)

Tornando alla PDI (l' **Informazione Descrittiva per la Conservazione**), è importante sottolineare che l'OAIS ne individua quattro categorie:

- ◆ **Identificazione** (Reference Information): enumera e descrive gli identificatori assegnati al Contenuto (p. es. l'URN di **1) Informazione sul Contenuto** e il riferimento alla normativa sull'URN)
- ◆ **Contesto** (Context Information): documenta le relazioni del Contenuto con il suo ambiente (perché è stato creato, in che relazione è con altri Contenuti) (p.es. altri O.I. nella stessa collezione, manifestazioni precedenti dello stesso Oggetto)
- ◆ **Provenienza** (Provenance Information): documenta la storia del Contenuto e i cambiamenti da esso subiti, oltre che la catena di custodia (p. es. il formato originale dei dati, quale specifico processo ha permesso la creazione /trasformazione di un determinato oggetto informativo, chi è stato responsabile di quel processo)
- ◆ **Autenticazione** (Fixity Information): documenta i meccanismi di autenticazione destinati ad assicurare l'integrità del Contenuto (p. es. l'impronta digitale calcolata con l'algoritmo MD5).

E' interessante notare che la definizione dei tipi di metadati necessari a comporre la PDI e' basata sulle proposte del Rapporto, del 1996, *Preserving Digital Information* della Task Force on Archiving of Digital Information, che per primo sottolineò come nell'ambiente digitale le categorie **Contenuto, Autenticazione, Identificazione, Provenienza e Contesto** fossero determinanti per il mantenimento dell'integrità e dell'autenticità dell'informazione e meritassero speciale attenzione per le finalità della conservazione a lungo termine.

Il Modello OAIS consiste quindi in una descrizione generale dei tipi di informazione richiesti dalle diverse funzioni di un sistema di archiviazione di dati. Proprio per questo, esso fornisce una cornice molto utile per sviluppare un set di metadati per la conservazione di larga applicabilità, come è del resto dimostrato dall'importanza delle iniziative che lo hanno adottato o che ad esso si riferiscono implicitamente.