

Un modello di dati per i metadati di conservazione

Nel giugno 2003 OCLC e RLG hanno creato un gruppo di lavoro internazionale sulle “strategie di implementazione dei metadati di conservazione” che avrebbe lavorato per due anni. Il Preservation Metadata: Implementation Strategies working group (PREMIS WG) era composto da 30 esperti, che rappresentavano le biblioteche, i musei, gli archivi, le agenzie di governo ed il settore privato, di 5 paesi diversi. L’obiettivo è stato quello di definire un “insieme di base di metadati di conservazione” per la comunità della conservazione digitale. Nel Maggio 2005 è stato pubblicato un rapporto conclusivo, che ha incluso un modello per i metadati di conservazione ed un dizionario nella versione 1.0. Al momento l’attività di implementazione è supportata dal gruppo di “maintenance” che sovrintende alla manutenzione dello schema e alle revisioni del dizionario dei dati.

Requisiti

Il PREMIS WG ha identificato le cinque aree a cui si riferiscono maggiormente i metadati di conservazione:

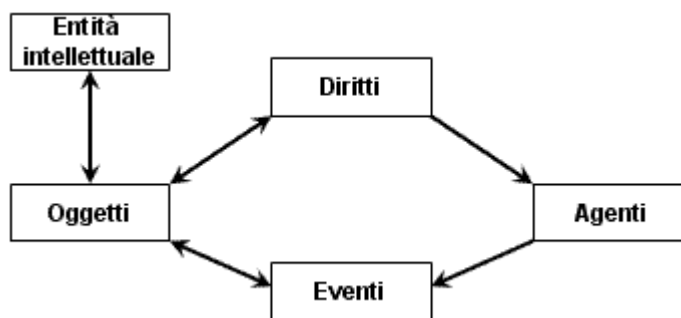
- Provenienza: dovrebbe raccogliere le informazioni riguardanti la storia della custodia dell’oggetto digitale, dal momento della sua creazione e andando avanti nei successivi cambiamenti di custodia e/o proprietà fisica.
- Autenticità: dovrebbe includere informazioni sufficienti ad assicurare che l’oggetto digitale archiviato sia quello che si presuppone, e che non sia stato alterato, intenzionalmente o non intenzionalmente, in modo non documentato.
- Attività di conservazione: dovrebbe documentare le azioni intraprese nel tempo per conservare l’oggetto digitale, e tenere memoria di qualsiasi conseguenza di tali azioni che impattano su qualsiasi caratteristica dell’oggetto.
- Ambiente tecnologico: dovrebbe descrivere hardware, sistema operativo ed applicazioni software, necessari a rendere fruibile l’oggetto digitale allo stesso modo in cui viene abitualmente conservato nel deposito.
- Gestione dei diritti: dovrebbe mantenere memoria di qualsiasi diritto di proprietà intellettuale vincolante, che limiti i poteri del deposito a intraprendere azioni per conservare l’oggetto digitale, e per abilitare la disseminazione agli utenti attuali e futuri.

Panoramica del Modello dei Dati

Il modello dei dati del PREMIS consiste di entità, relazioni e proprietà, che vengono chiamate unità.

Entità

- Entità Intellettuale – un insieme coerente di contenuto che può essere ragionevolmente descritto come un’unità, per esempio, un libro, una mappa, una fotografia o un database. Dal momento che questa entità è già ben descritta dai metadati descrittivi, il dizionario dei dati non li include.
- Oggetto o Oggetto Digitale – un’unità discreta di informazione nella forma digitale.
- Evento – un’azione che implica almeno un oggetto o un agente noto al deposito di conservazione.



- Agente – una persona, un’organizzazione, o un programma software associato agli eventi di conservazione nella vita di un oggetto.
- Diritti – asserzione di uno o più diritti o dei permessi legati ad un oggetto e/o ad un agente.

Relazioni

Le relazioni dichiarano le associazioni tra le istanze delle entità. Le “Relazioni” possono essere interpretate in modo ampio o ristretto, e qualsiasi fatto relazionale può essere espresso in molti modi diversi. Le Relazioni tra gli Oggetti possono essere di tre tipi:

- Le relazioni strutturali mostrano le relazioni tra le parti degli oggetti. Le relazioni strutturali tra i *file*, che costituiscono una rappresentazione di un’entità intellettuale, sono chiaramente essenziali come metadati di conservazione. Se un deposito non può ricostruire le varie parti di un oggetto digitale, si può affermare che non ha conservato l’oggetto.

Further information and resources

PREMIS (PREservation Metadata: Implementation Strategies) Resources.

<http://www.oclc.org/research/projects/pmwg/resources.htm>

Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group.

<http://www.oclc.org/research/projects/pmwg/premis-final.pdf>

Brian Lavoie, Richard Gartner, Preservation Metadata. DPC Technology Watch Report No. 05-01: September 2005

<http://www.dpconline.org/docs/reports/dpctw05-01.pdf>

PREMIS Working Group.

<http://www.oclc.org/research/projects/pmwg/>

PREMIS Maintenance Activity.

<http://www.loc.gov/standards/premis/>

- Le relazioni di derivazione risultano dalla replicazione o dalla trasformazione di un Oggetto. Il contenuto intellettuale dell'Oggetto di risulta è lo stesso, ma le istanze dell'oggetto e probabilmente il suo formato, sono diversi. Molti oggetti digitali sono complessi e sia le informazioni strutturali che di derivazione, possono cambiare nel tempo come risultato delle attività di conservazione.

- La relazione di dipendenza esiste laddove un oggetto ne richieda un altro a supporto della sua funzione, o della sua distribuzione o della coerenza del contenuto. L'oggetto di supporto formalmente potrebbe non essere parte dell'oggetto stesso, ma è necessario alla sua rappresentazione.

Proprietà

Le unità semantiche sono le proprietà di un'entità. In alcuni casi un'unità semantica può essere un contenitore che raggruppa un insieme di unità semantiche correlate e le sottounità raggruppate sono chiamate componenti semantici dell'unità semantica.

Il principio 1:1

Il principio 1:1 nei metadati asserisce che ogni descrizione faccia riferimento ad una ed una sola risorsa. Applicato ai metadati PREMIS, ogni Oggetto (*file*, *bitstream*, rappresentazione), tenuto in un deposito di conservazione viene descritto come uno statico insieme di bit. Non è possibile modificare questo insieme, ma solo creane uno nuovo, che è correlato all'Oggetto originario con una relazione di derivazione. Infatti, il Dizionario dei Dati contiene un'unità semantica solo per la data di creazione di un Oggetto, in quanto per definizione, non può essere modificato.

Dal modello al dizionario

Al modello dei dati viene associato un dizionario che comprende tutte le unità semantiche pertinenti e che descrivono le quattro entità (Oggetti, Agenti, Eventi, Diritti). Nel modello dei dati, le Relazioni tra le Entità sono messe in evidenza dalle frecce. Il Dizionario dei Dati le esprime come informazioni di collegamento, includendo nelle informazioni dell'entità A un puntatore all'entità B correlata. Ogni entità del modello dei dati ha un identificatore univoco usato come puntatore. Per esempio, l'entità Oggetto contiene frecce che puntano alle Entità Intellettuali ed agli Eventi. Queste sono implementate nel Dizionario dei Dati, dalle unità semantiche `linkingIntellectualEntityIdentifier` e `linkingEventIdentifier`.

Pro e Contro

Il dizionario dei dati del PREMIS è il risultato di un processo di costruzione di consenso internazionale e multidisciplinare, che può accrescere ampiamente le opportunità di applicazione nei diversi tipi di istituzioni, scenari della conservazione digitale e sistemi di implementazione. Come schema di metadati di conservazione, tenta di sopperire al massimo alle attuali necessità finalizzate all'implementazione pratica e all'interoperabilità e facilita le transazioni degli oggetti, nell'assolvere alla sua funzione principale: documentare gli oggetti digitali nel tempo, rendendoli accessibili a lungo termine. Il lato oscuro del dizionario PREMIS è che, nessuno può dire con certezza se la sua efficacia sarà durevole. L'impatto degli sviluppi futuri, negli usi degli oggetti, rimane difficile da prevedere.

Conclusioni

Il dizionario dei dati costituisce una parte cruciale dell'infrastruttura di conservazione digitale, ed è il componente essenziale con cui possono essere implementate strategie di conservazione digitale efficaci e sostenibili.