



# Lo standard PREMIS e la conservazione digitale

## modulo di base – I parte

Angela Di Iorio



Quest'opera è distribuita con  
Licenza [Creative Commons](https://creativecommons.org/licenses/by-nc-sa/4.0/)  
[Attribuzione - Non commerciale -](https://creativecommons.org/licenses/by-nc-sa/4.0/)  
[Condividi allo stesso modo 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)  
[Internazionale.](https://creativecommons.org/licenses/by-nc-sa/4.0/)



# Lo standard PREMIS e la conservazione digitale

Roma, 9-11 Giugno 2016

Angela Di Iorio



SAPIENZA  
UNIVERSITÀ DI ROMA

<http://sbs.uniroma1.it/sdl/premis>

[angeladiiorio@gmail.com](mailto:angeladiiorio@gmail.com)

[angela.diiorio@uniroma1.it](mailto:angela.diiorio@uniroma1.it)

alcuni materiali di questo corso sono stati estratti e tradotti dalla documentazione e dai materiali di supporto reperibili nel sito web di PREMIS

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

...e ricontestualizzati per gli obiettivi e la durata di questo seminario



—  
Quest'opera è distribuita con  
Licenza [Creative Commons](https://creativecommons.org/licenses/by-nc-sa/4.0/)  
[Attribuzione - Non commerciale -](https://creativecommons.org/licenses/by-nc-sa/4.0/)  
[Condividi allo stesso modo 4.0](https://creativecommons.org/licenses/by-nc-sa/4.0/)  
[Internazionale.](https://creativecommons.org/licenses/by-nc-sa/4.0/)

## ✦ Introduzione, modello dei dati e strumenti di supporto

- ✦ Contesto di applicazione e la storia breve dello sviluppo di PREMIS
- ✦ Lo standard di riferimento
- ✦ Il modello dei dati: le entità e le relazioni tra entità  
(versione 2 e differenze con la versione 3)
- ✦ Gli strumenti di supporto
  - ✦ Il dizionario dei dati e le unità semantiche
  - ✦ Il sito web ufficiale

## ✦ **Mantenere la conformità allo standard** **- entità “Oggetto” e relazioni con le altre entità**

- ✦ Linee guida di applicazione
- ✦ Livelli di conformità
- ✦ L'entità **Oggetto** elementi obbligatori
- ✦ Approcci di implementazione e problemi
- ✦ Casi d'uso per applicare PREMIS
- ✦ Evoluzione di PREMIS

## ✦ **Mantenere la conformità allo standard - entità “Agente”, “Evento”, “Diritti”**

✦ Le entità Agente, Evento, Diritti

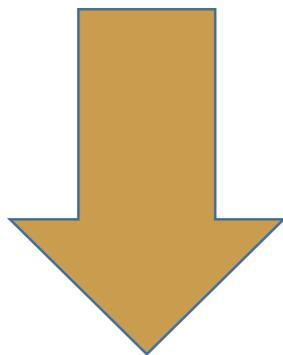
✦ Discussione e *Question Time*

✦ Conclusioni

## ⊠ Argomenti supplementari per esercitazione e casi d'uso

- ⊠ Approcci di implementazione e problemi
- ⊠ Casi d'uso per applicare PREMIS
- ⊠ Evoluzione di PREMIS

✦ **Contesto di applicazione**



**Conservazione Digitale  
a Lungo Termine**



▣ Contesto di applicazione

## Qual è la relazione con la Conservazione Digitale a Lungo Termine?

- L'esistenza di un oggetto digitale dipende dalla sua corretta conservazione digitale.
- La corretta conservazione digitale inizia nel momento della **nascita digitale** dell'oggetto.
- La crescita di consapevolezza della comunità digitale, in merito alla conservazione, si è manifestata dalle iniziative di implementazione. Tali iniziative hanno messo in evidenza il bisogno di una infrastruttura di metadati per la conservazione digitale.

Contesto di applicazione

## Metadati di conservazione

- **Provenienza**
  - Chi ha la custodia/proprietà dell'oggetto digitale?
- **Autenticità**
  - L'oggetto digitale è quello che si presuppone essere?
- **Attività di conservazione**
  - Cosa è stato fatto per preservarlo?
- **Ambiente tecnologico**
  - Cosa è necessario per essere restituito dall'ambiente digitale (render) e quindi eventualmente, riusarlo?
- **Informazione strutturale**
  - Come sono messi insieme gli oggetti digitali?
- **Gestione dei diritti**
  - Quali vincoli legali devono essere osservati?



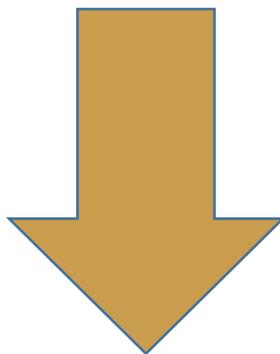
- 10 anni

- 50 anni

- futuro indistinto

**Rendere gli oggetti digitali auto-documentativi (self-documenting) nel tempo**

## ✦ Storia breve di PREMIS



**Bisogni emersi dalla breve  
storia del mondo digitale**

## Perchè PREMIS

**Standard internazionale per i metadati di supporto alla conservazione degli oggetti digitali che permette di rendere persistente la loro usabilità.**

- Come implementare un Sistema di deposito digitale deputato alla conservazione degli oggetti digitali?
  - Quali sono i metadati minimi necessari a garantire l'accesso agli oggetti digitali?
  - Quali sono le entità informative coinvolte e come si relazionano?

### *PREMIS: Preservation Metadata: Implementation Strategies*

- Sviluppato da un team internazionale di esperti (the PREMIS Working Group)
- Implementato nei progetti di conservazione digitale in diversi paesi intorno al globo
- Usato da sistemi e applicazioni commerciali e open-source per la conservazione digitale



## Storia breve

**Pre-2002:** diversi insiemi di metadati di conservazione sono stati rilasciati

- Iniziative diverse, obiettivi diversi, ambito diverso e diversi i modelli e i presupposti di base
- Nessuno standard internazionale; nessuna buona pratica e competenze di settore poco consolidate

**Giugno 2002:** Preservation Metadata Framework

- Gruppo di lavoro Internazionale, sponsorizzato da OCLC e RLG
- Metadati di conservazione consistenti in una descrizione ad alto livello delle tipologie previste
- Ispirato dal modello di riferimento Open Archive Information System (OAIS) come punto di partenza
- Prototipo di insieme degli elementi di metadati di conservazione
- Una soluzione basata sul consenso per lo sviluppo di specifiche formali per i metadati di conservazione ... ma non una soluzione pronta per l'implementazione ("off-the-shelf, ready to implement"), la quale implicherebbe una dipendenza tecnologica

**Post-2002:** Bisogno di metadati di conservazione implementabili, con accluse linee guida per l'applicazione e l'uso, e pertinenti per un ampio ventaglio di sistemi di conservazione digitale e di contesti applicativi

- Formazione motivata del **PREMIS Working Group**



☒ Storia breve

## PREMIS Working group

- **Giugno 2003:** OCLC, RLG ha sponsorizzato un nuovo gruppo di lavoro internazionale:
  - *PREMIS: Preservation Metadata: Implementation Strategies*
- **Composto da:**
  - **più di 30 esperti di 5 paesi, rappresentanti di biblioteche, musei, archivi, agenzie governative, e del settore privato**
  - Co-Chairs: Priscilla Caplan (FCLA), Rebecca Guenther (LC)
- **Objective 1:** Identificare e valutare strategie alternative per codificare, archiviare gestire e scambiare metadati di conservazione
  - PREMIS Survey Report (September 2004)
- **Objective 2:** Definire il cuore di metadati implementabili di conservazione, con linee guida e raccomandazioni per la gestione e l'uso

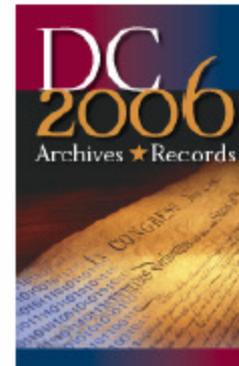


## I premi assegnati a PREMIS

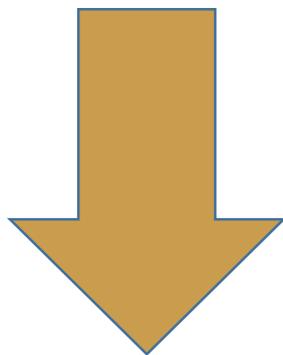
### 2005 British Conservation Awards: Digital Preservation Award



### 2006 Society of American Archivists Preservation Publication Award



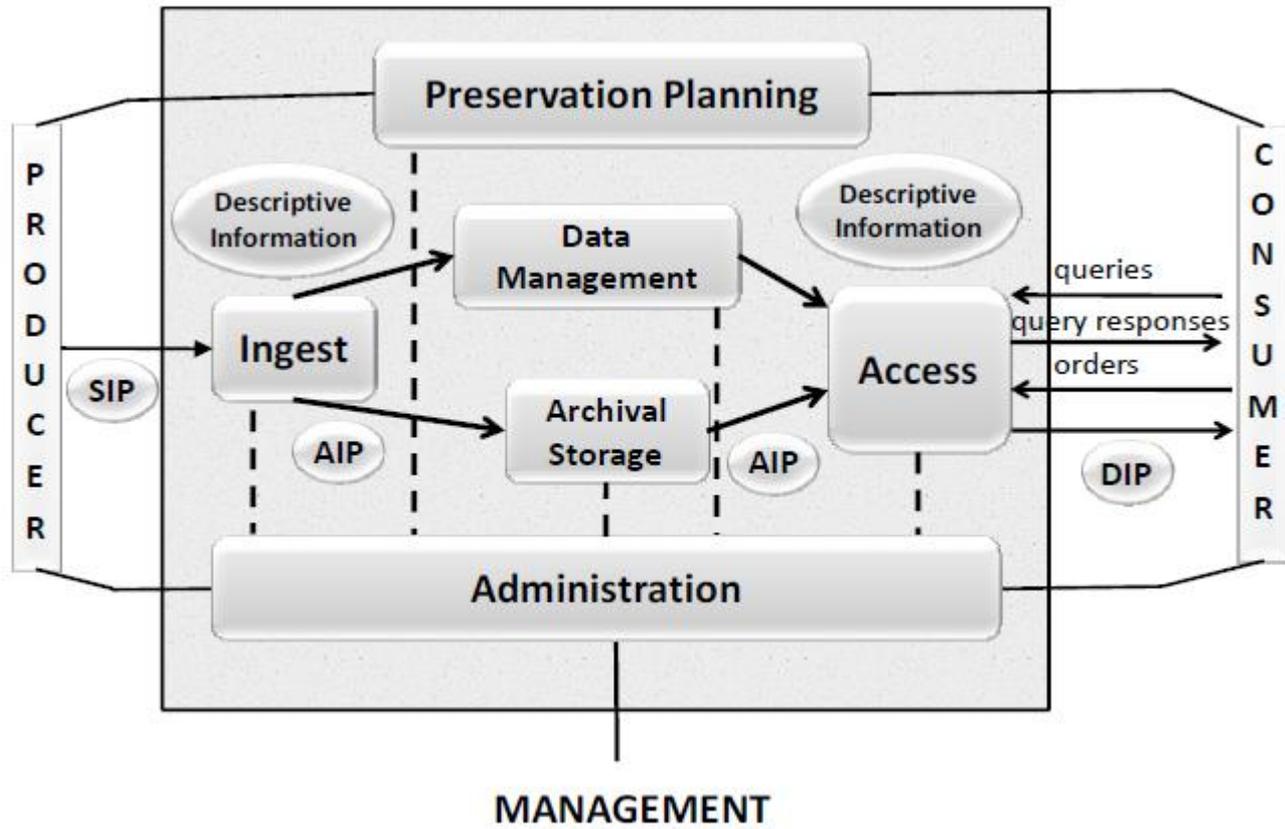
## 3 – Lo standard di riferimento



**Open Archival Information System  
(OAIS)**

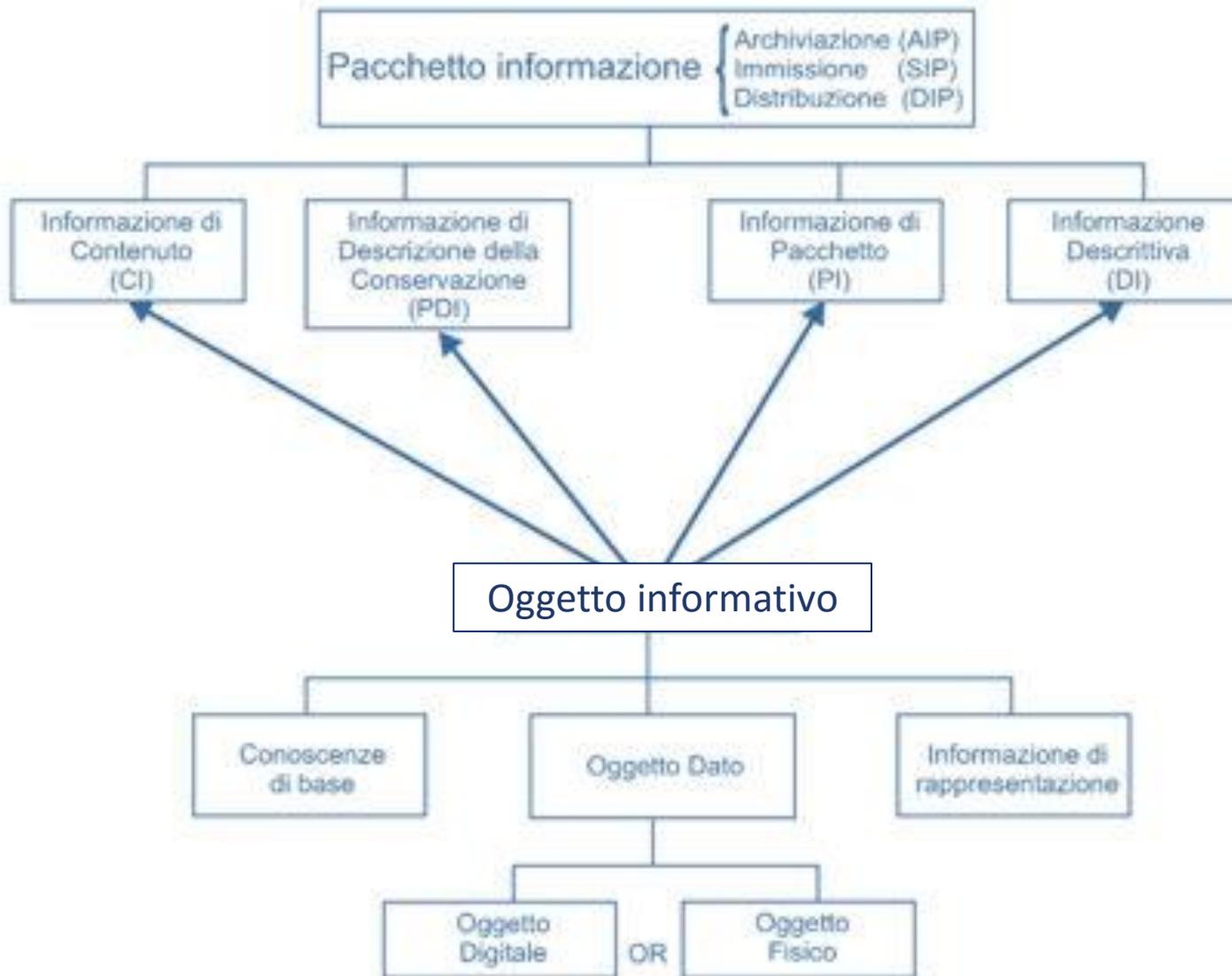
## Standard di riferimento

- Lo standard di riferimento Open Archive Information System



Consultative Committee for Space Data Systems  
Reference Model for an Open Archival Information  
System (OAIS), Recommended Practice CCSDS 650.0-  
M-2 Magenta Book, June 2012,  
<http://public.ccsds.org/publications/archive/652x0m1.pdf>

➔ Standard di riferimento – OAIS 2002



Standard di riferimento – OAIS 2002



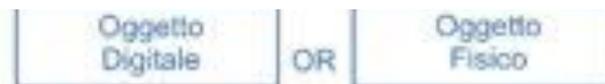
Il pacchetto informativo (IP) è un'aggregazione di queste quattro classi di oggetti.

Gli IP possono essere di tre tipi:

**SIP - (Submission Information Package)** - *IP di Immissione*, riversato dal produttore nell'archivio.

**AIP - (Archival Information Package)** - *IP di Archiviazione*, che è il pacchetto conservato nell'archivio.

**DIP - (Dissemination Information Package)** - *IP di Distribuzione* trasferito dall'archivio all'utente in risposta ad una richiesta di accesso.





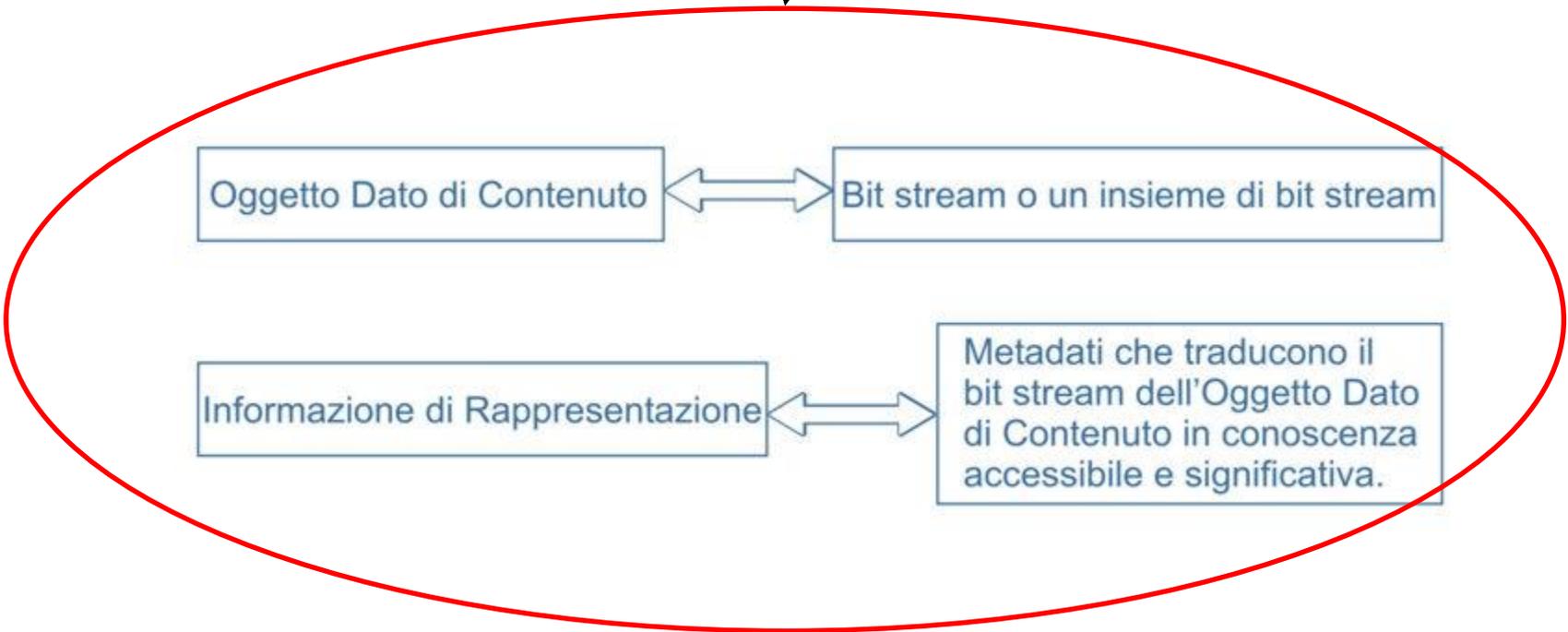
Standard di riferimento – OAIS 2012

## OAIS - Tipologie di “Pacchetto di Informazione” (IP)

- Submission Information Package (SIP) **IP di sottomissione**
  - Caricato (Ingestion) in un deposito digitale
  - Potrebbe includere metadata del produttore/distributore
  - Informazione potrebbe essere estratta per essere usata nella struttura di archiviazione del deposito
- Archival Information Package (AIP) **IP di Archiviazione**
  - Oggetto di conservazione, il COSA viene archiviato
  - Conforme ai requisiti del deposito
  - Include metadati o collegamenti ai metadati che supportano la conservazione digitale
- Dissemination information package (DIP) **IP di Disseminazione**
  - Usato per distribuire contenuti e metadati
  - Cosa viene acceduto dalla comunità designata.

↩ **Standard di riferimento - OAIS**

**Oggetto di conservazione**





## Standard di riferimento – OAIS 2002

**AIP - (Archival Information Package)** - *IP di Archiviazione*, che è il pacchetto conservato nell'archivio.

**Informazione di Contenuto (CI)** - contiene l'Oggetto Dati di Contenuto che viene affidato all'archivio con le sue informazioni di rappresentazione associate

**Informazione di descrizione della conservazione (PDI)** - contiene l'informazione necessaria alla gestione della conservazione del CI a cui è associata.

Il modello OAIS identifica quattro tipi di PDI:

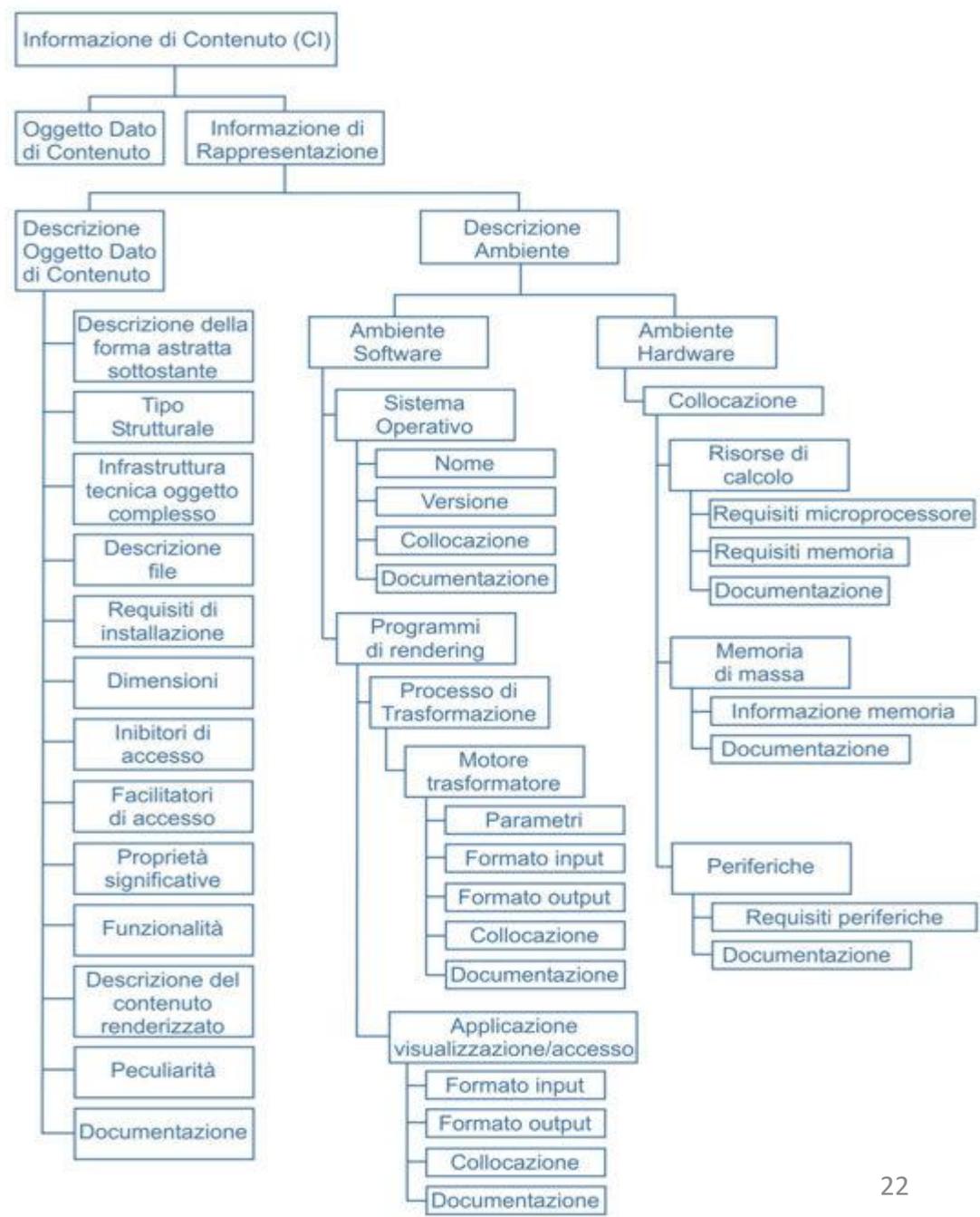
- *Informazione di Riferimento*, enumera e descrive gli identificatori assegnati al CI come quelle che possono essere riferite inequivocabilmente agli archivi, sia internamente che esternamente, (p.e., ISBN, URN)
- *Informazione di Provenienza*, documenta la storia del CI (ad esempio, origini, interventi ed effetti di conservazione e di tutela)
- *Informazione di Contesto*, documenta le relazioni del CI con il suo ambiente (ad esempio, perchè è stato creato e quali sono le relazioni con altre CI)
- *Informazione di Persistenza*, documenta i meccanismi di autenticazione utilizzati per assicurare che il CI non sia stato alterato in modo non documentato (ad esempio, firma digitale, checksum)

**Informazione di Pacchetto (PI)** - raccoglie l'oggetto digitale e i suoi metadati associati in una unità o pacchetto identificabile (ad esempio, in un AIP)

**Informazione Descrittiva (DI)** - facilita l'accesso alla CI con strumenti di ricerca e recupero dell'informazione, dall'archivio. Tipicamente viene derivata dalla CI e dalla PDI.

Standard di riferimento  
- OAIS -  
Informazione di  
Contenuto

**Informazione di Contenuto (CI)**  
contiene l'Oggetto Dati di  
Contenuto che viene affidato  
all'archivio con le sue informazioni  
di rappresentazione associate





## Standard di riferimento – OAIS 2002

**AIP - (Archival Information Package)** - *IP di Archiviazione*, che è il pacchetto conservato nell'archivio.

### Informazione di descrizione della conservazione (PDI)

**Informazione di Riferimento:** descrive i sistemi di identificazione, ed il meccanismo per fornire gli identificatori assegnati, che vengono usati per identificare inequivocabilmente la CI, sia internamente che esternamente all'archivio in cui è contenuto.

**Informazione di Contesto:** relazioni dei documenti della CI con il suo ambiente. Include le ragioni della sua creazione e le relazioni con altri oggetti di CI.

**Informazione di Provenienza:** la storia dei documenti della CI. Include le sue origini, i cambiamenti fatti sull'oggetto e sul suo contenuto nel corso del tempo, e infine sulla sua catena di custodia.

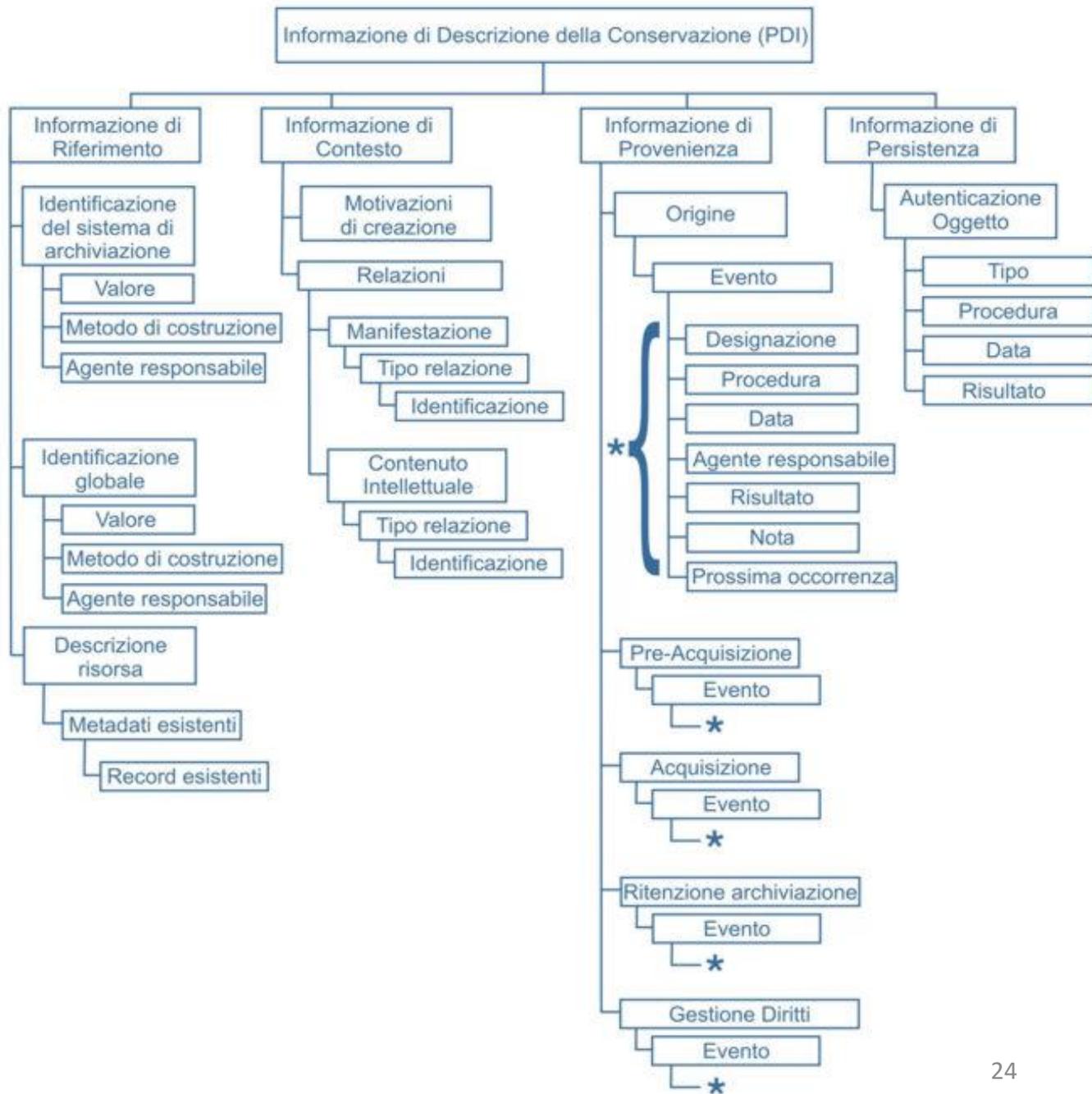
**Informazione di Persistenza:** fornisce il controllo dell'integrità dei dati o le chiavi di validazione e verifica. Viene usata per assicurare che il particolare oggetto della CI non abbia subito alterazioni in modo non documentato.

Per riassumere la PDI registra le informazioni dell'Oggetto Dato di Contenuto archiviato, in merito a:

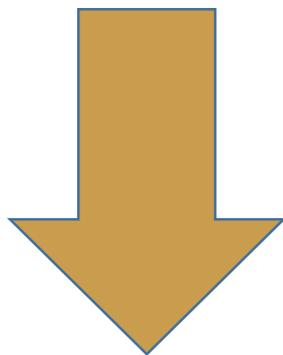
**Identità, Relazioni, Storia e Integrità.**

## ⊕ - OAIS - Informazione di Descrizione della Conservazione

**Informazione di descrizione della conservazione (PDI)** contiene l'informazione necessaria alla gestione della conservazione della CI a cui è associata.



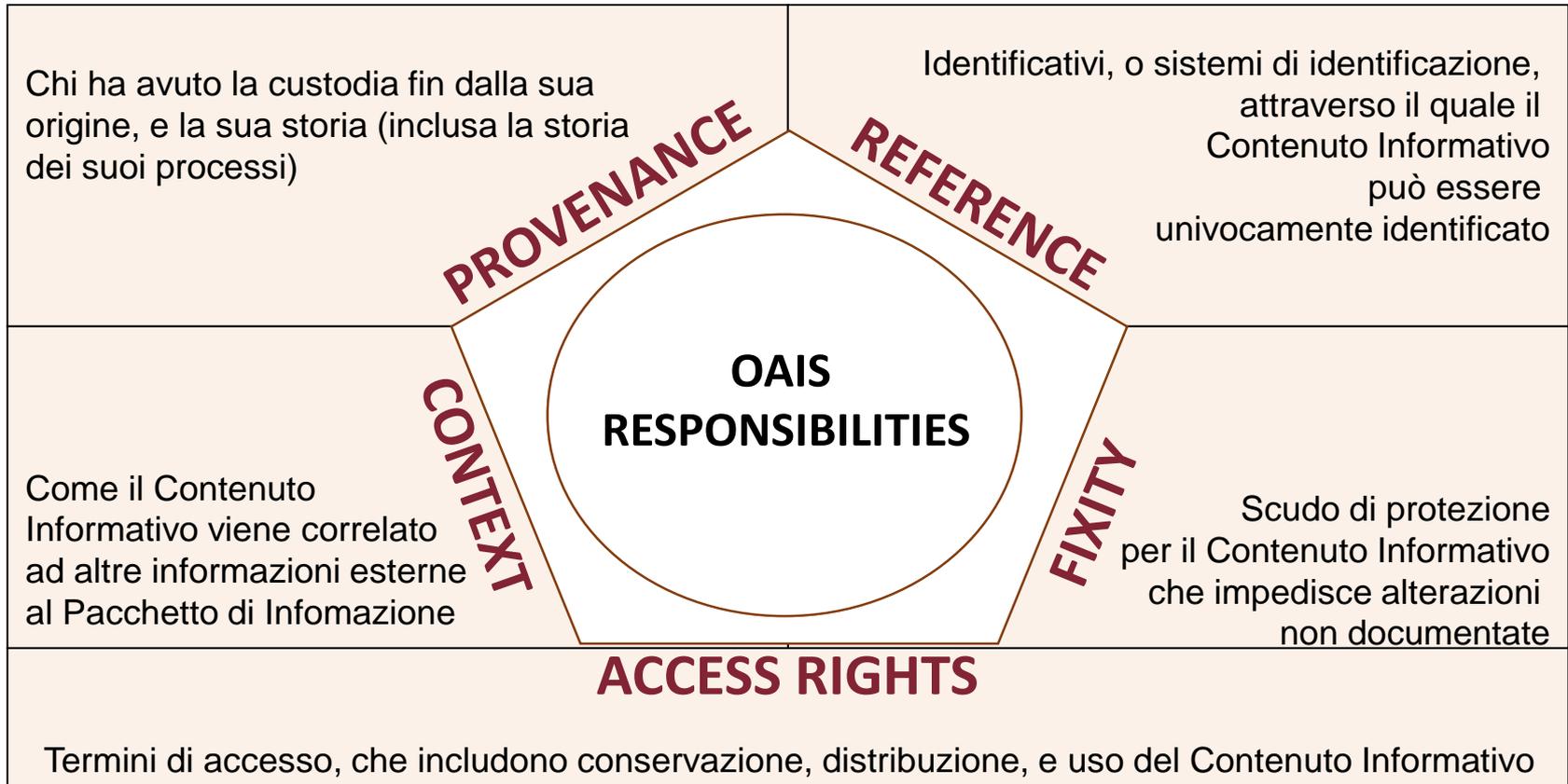
## ✦ OAIS e conservazione



**Dati necessari e perchè**



Standard di riferimento OAIS 2012



↳ **Obiettivi della Conservazione Digitale**



Priscilla Caplan (FCLA)



## ☒ Trusted digital repository: funzioni che usano i metadati di conservazione

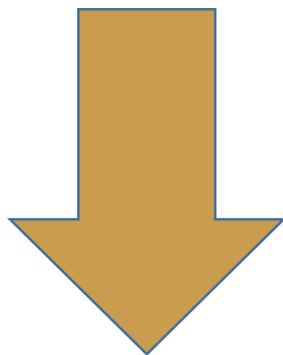
- Identificare le proprietà da conservare
- Verificare ogni oggetto che viene sottoposto all'acquisizione del deposito (*Ingestion*)
- Creare il pacchetto di archiviazione dal pacchetto di *ingestion* per includere metadati tecnici e di uso legale
- Disporre di meccanismi per autenticare il contenuto e la sua fonte
- Disporre di meccanismi per assicurare che il contenuto non venga corrotto
- Gestire le copie con le relative collocazioni degli oggetti digitali
- Mantenere descrizioni precise delle azioni di conservazione per assicurare che gli oggetti vengano preservati
- Usare strumenti e risorse come i registri dei formati per rilevare i formati caduti in obsolescenza.
- Tracciare e gestire diritti e restrizioni



✚ Trusted digital repository:  
funzioni che usano i metadati di conservazione

- **Fixity** - Checksum archiviato come metadato in modo da permettere al Sistema di controllare che gli oggetti non siano stati alterati
- **Media management** - informazione riguardante i dispositivi fisici dove i file vengono archiviati: tipologie, età dei dispositivi e date in cui sono stati controllati (*refreshment*)
- **Format obsolescence** - controllo dei formati dei file e sviluppo delle strategie necessarie a mantenerli in uso - l'informazione riguardante 1) *formato originale*, e ambiente 2) *software* e 3) *hardware* è necessaria per supportare i formati obsoleti
- **Digital provenance** - informazione necessaria a tracciare la catena di custodia e la storia dei cambiamenti

## ✦ Il modello dei Dati



**Il modello dei dati (versione 2/3)**

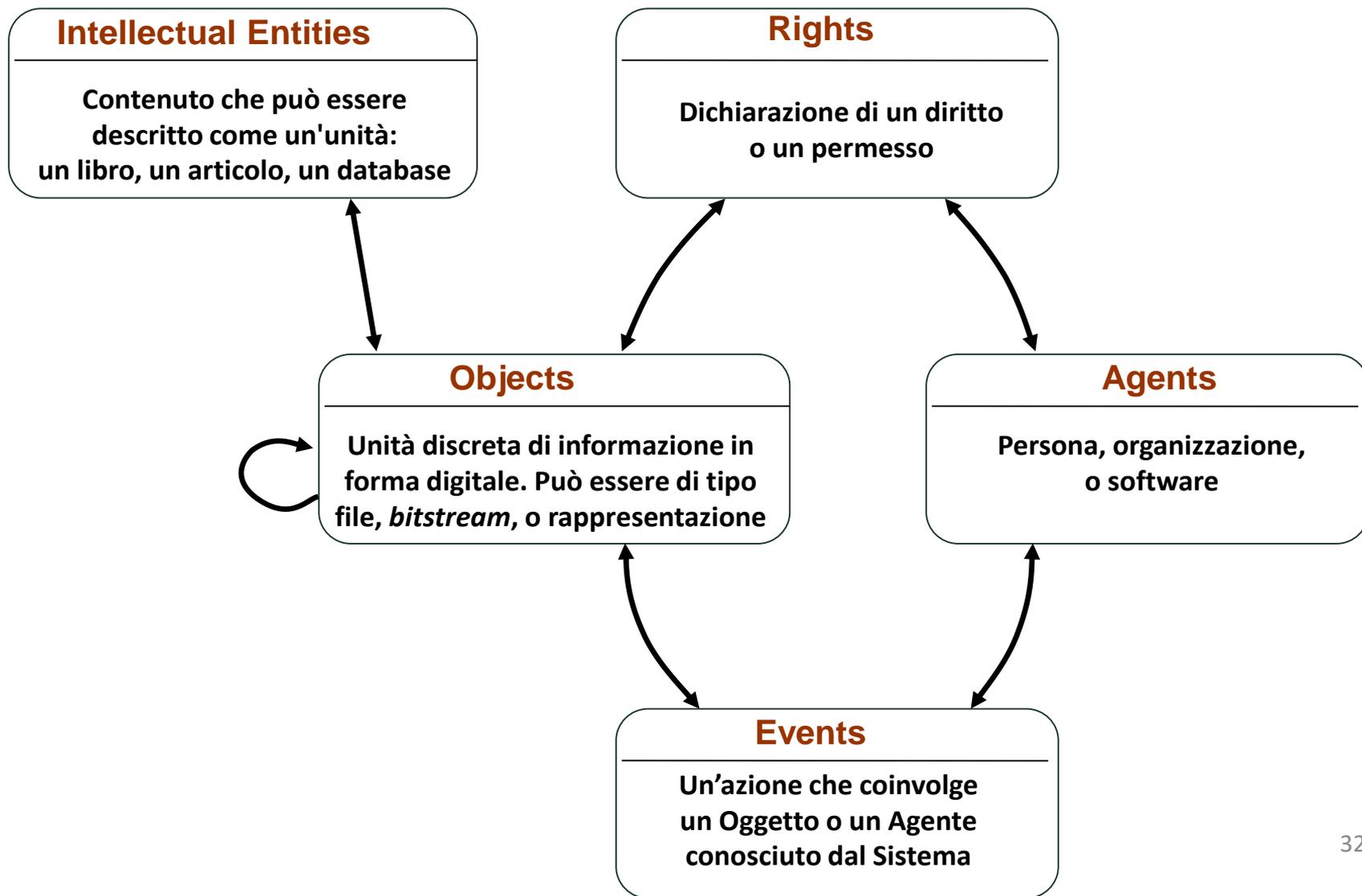


## Il modello dei dati: perchè

- **Facilitatore di apprendimento per la comunità di riferimento:** permette di comprendere più agevolmente la complessità descrittiva e di individuare le entità di riferimento
- **Facilitatore per lo sviluppo di strategie di conservazione:** è conveniente per un'organizzazione allo scopo di comprendere come sviluppare la conservazione dei propri beni digitali
- **Collettore di informazione per entità distinte:** la raccolta dei dati intorno a specifiche entità permette di distinguere la competenza dei dati ad un tipo di oggetto piuttosto che ad un altro

## Il modello dei dati fino versione 2.x

*Understanding PREMIS* by Priscilla Caplan for the Library of Congress,  
<http://www.loc.gov/standards/premis/understanding-premis.pdf> p. 8



## Intellectual Entity

### Intellectual Entities

Contenuto che può essere descritto come un'unità:  
un libro, un articolo, un database

**Intellectual Entity:** un insieme di contenuto che viene considerato come una singola entità intellettuale per scopi di gestione e di descrizione: per esempio, un libro particolare, una mappa, una fotografia, o un database. Può includere altre Entità Intellettuali: ad esempio, un sito web che include una pagina web, che include un'immagine web.

Un'Entità Intellettuale potrebbe avere una o più rappresentazioni digitali.

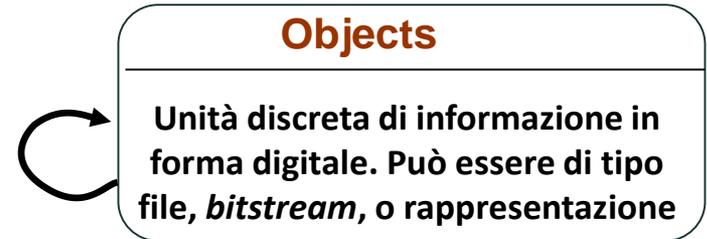
Esempi:

- *The Chamber* by John Grisham(an ebook)
- “Maggie at the beach” (a photograph)
- The Library of Congress Website (a website)

Inizialmente, non veniva descritta in modo esaustivo nel PREMIS DD, e poteva essere collegata con altri metadati che ne descrivono la sua rappresentazione digitale ... **ciò è stato affrontato nella versione 3.0**

## Entità Oggetto

**Object (or Digital Object):** Unità discreta di informazione in forma digitale



Gli *Objects* sono ciò che effettivamente viene immagazzinato e gestito dal sistema di deposito digitale. La maggior parte dei metadati di PREMIS sono incentrati sulla descrizione degli oggetti digitali.

L'entità Oggetto può essere di tre sottotipi: **file**, **bitstream**, **representation** (versione 3.0 include l'**entità intellettuale di tipo ambiente**)

Un **file** è una sequenza di bytes ordinata e specificamente denominata che è conosciuta al sistema operativo. Un file potrebbe essere di 0 o più bytes ed ha un formato di file, permessi di accesso, e caratteristiche proprie del file system come le dimensioni(size) e l'ultima data di modifica.

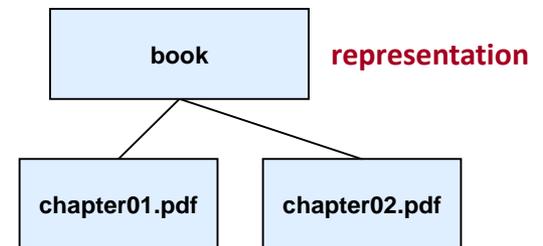
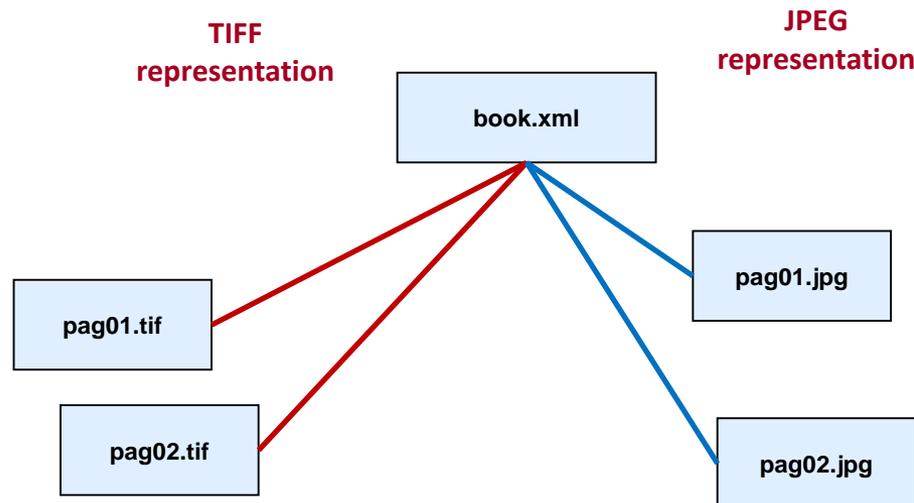
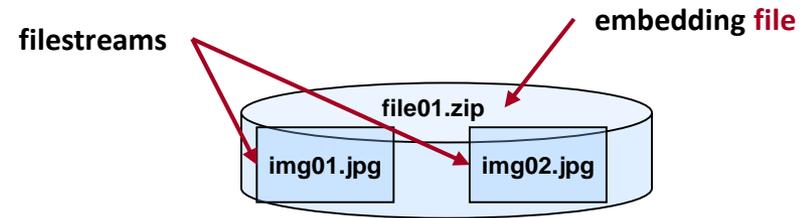
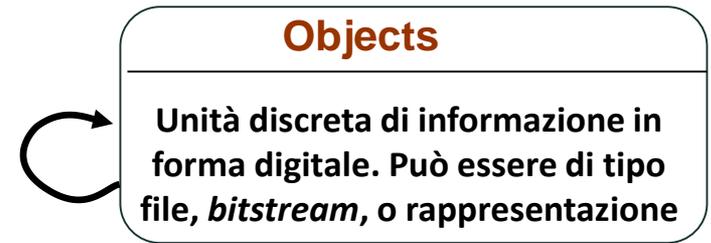
Un **bitstream** è costituito da dati contigui o non-contigui all'interno di un file che ha proprietà significative e comuni per gli scopi della conservazione. Non può essere trasformato in un file a se stante senza l'aggiunta di una struttura di file (headers, ecc.) e/o la sua riformattazione per essere conforme ad uno specifico formato di file.

Una **representation** è l'insieme dei metadati relativi ai file e al come sono messi insieme (metadati strutturali), e sono necessari alla completa traduzione (rendition) di un'Entità Intellettuale. Ad esempio, un articolo di una rivista potrebbe essere completo in un file PDF; questo singolo file costituisce anche la rappresentazione della Entità Intellettuale. Un altro articolo di rivista potrebbe consistere in un file SGML e due file immagine; questi tre file ne costituiscono la sua rappresentazione. Un terzo articolo potrebbe essere costituito da 12 pagine in formato TIFF (1 file per ogni pagina) ed un file XML dove i metadati strutturali associano il numero delle pagine ai file TIFF; la rappresentazione è costituita da 13 file.

## Object entity

Esempi:

- un file PDF
- un libro composto da diversi file XML e molte immagini anche di formato diverso
- un'immagine in un file PDF
- un file compresso





## Entità Agente

### Agents

Persona, organizzazione,  
o software

**Agent:** persona, organizzazione, o software (programma/ssistema) associato con gli Eventi nel ciclo di vita di un Oggetto, o con i Diritti (Rights) (dichiarazione di diritto a...) che hanno un qualche impatto sull'Oggetto.

- Gli Agenti sono associati indirettamente agli Oggetti, ovvero attraverso gli Eventi o i Diritti
- Non vengono descritti in dettaglio nel PREMIS DD perchè non sono considerati metadati di conservazione "core", ma la loro specifica identificazione dovrebbe rendersi necessaria.

### Examples:

- Angela Di Iorio (a person)
- Università di Roma la Sapienza (an organization)
- Massive Metadata Production System (a system)
- BRI-DGE version 2.0 (a software program)



⊞ Event entity

## Events

Un'azione che coinvolge un Oggetto o un Agente conosciuto dal Sistema

**Event:** Un'azione che coinvolge un Oggetto o un Agente conosciuto dal Sistema di conservazione digitale.

- Documenta la provenienza digitale. Può tracciare la storia di un Oggetto attraverso la catena di Eventi che occorrono durante il ciclo di vita degli Oggetti.
- Determina quali sono gli Eventi di competenza del deposito ( eventualmente anche Eventi di pre o post gestione )
- Determina quali sono gli Eventi che dovrebbero essere memorizzati, e a quale livello di granularità (viene deciso dal sistema di deposito stesso)

### Esempi:

- Evento di Validazione: uso di JHOVE per verificare che il file PDF chapter01.pdf sia effettivamente un file PDF valido
- Evento Ingest: trasformazione di un OAIS SIP in un AIP (uno o più Eventi?)
- Evento migrazione: creare una nuova versione di un Oggetto in un formato più aggiornato



## Entità Rights

### Rights

Dichiarazione di un diritto  
o un permesso

**Rights:** dichiarazione di un diritto o di un permesso che si riferisce ad un Oggetto e/o ad un Agente.

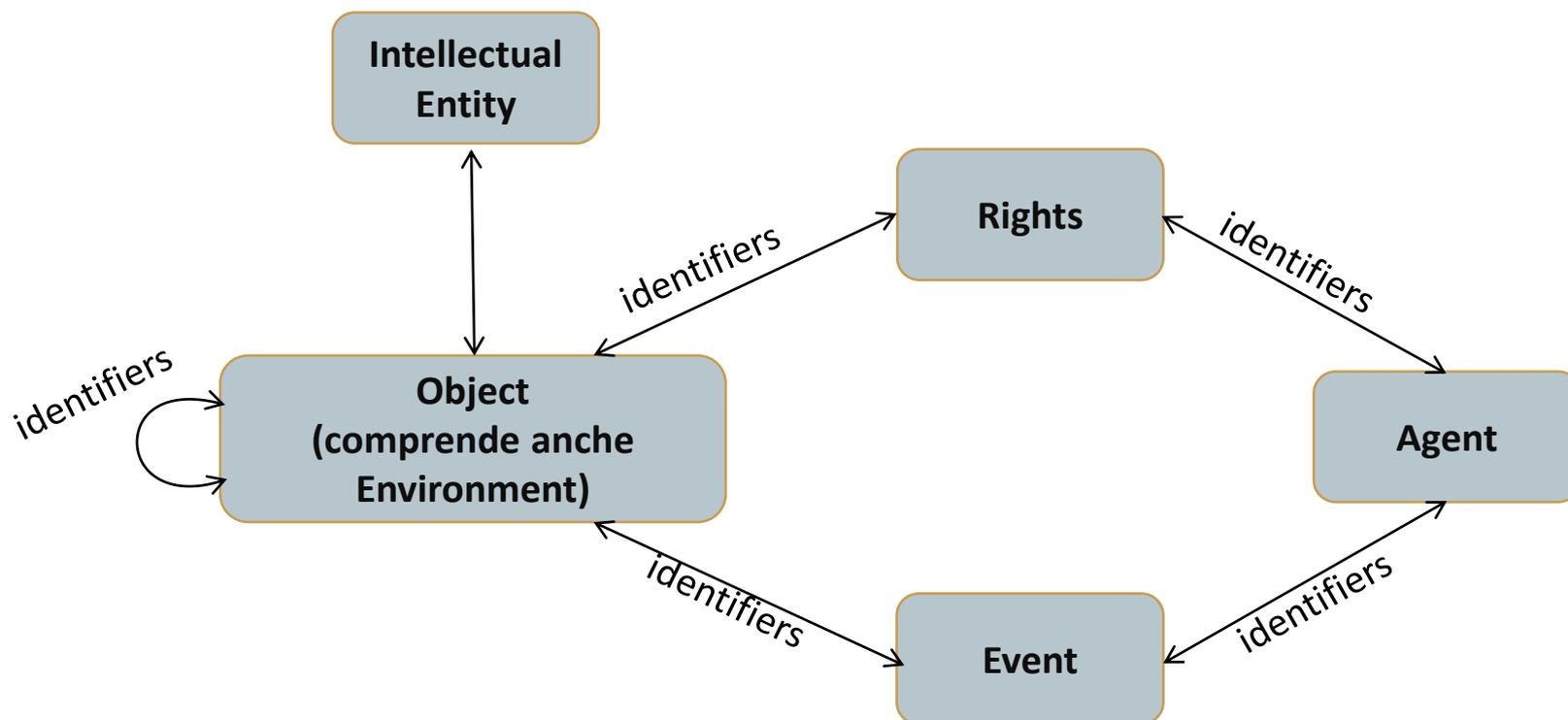
- Un accordo con chi "detiene i diritti", il quale garantisca il permesso al deposito di intraprendere le azioni necessarie a preservare gli oggetti del deposito.
- Non è un linguaggio completo che esprime diritti; è incentrato esclusivamente sui permessi che assumono la seguente forma:  
Agente X ha il permesso (Rights) Y nel deposito in riferimento all' Oggetto Z.
- I Diritti potrebbero essere associati attraverso diverse forme che vedremo in dettaglio

### Esempio:

Il proprietario dell'Oggetto Z, assegna al sistema di archiviazione X il permesso Y di fare tre copie dell'Oggetto Z, a scopo di conservazione.



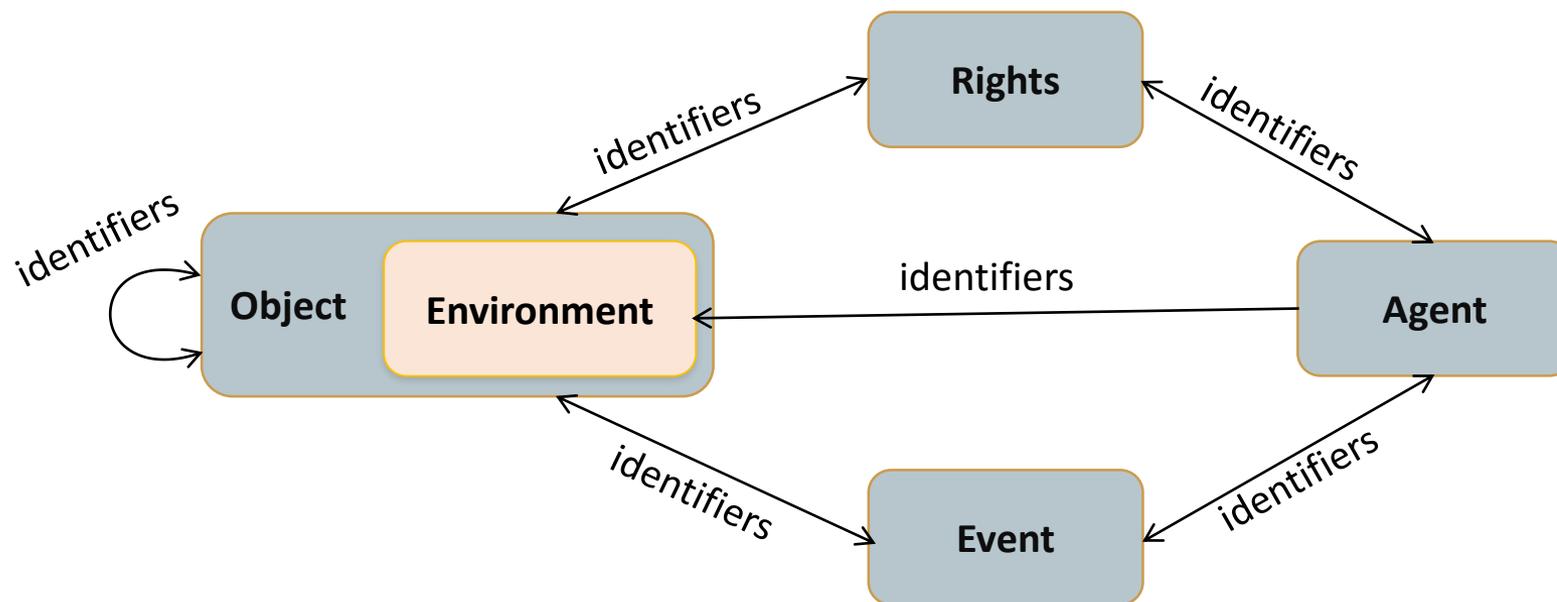
Il modello dei dati fino versione 2.x



Il principale cambiamento dalla versione 2 alla versione 3 è principalmente focalizzato sull'ambiente tecnologico di supporto dell'oggetto:

fino alla versione 2, l'ambiente viene descritto semplicemente come unità semantica appartenente all'Oggetto, senza sottolinearne le relazioni di dipendenza e senza avere una dignità come Oggetto a se stante.

## Il modello dei dati Versione 3.0



Il principale cambiamento dalla versione 2 alla versione 3 è principalmente focalizzato sull'ambiente tecnologico di supporto dell'oggetto:

nella versione 3 l'ambiente può essere descritto in modo più esaustivo assumendo anche la dignità di Oggetto a se stante da cui dipendono gli oggetti di contenuto.

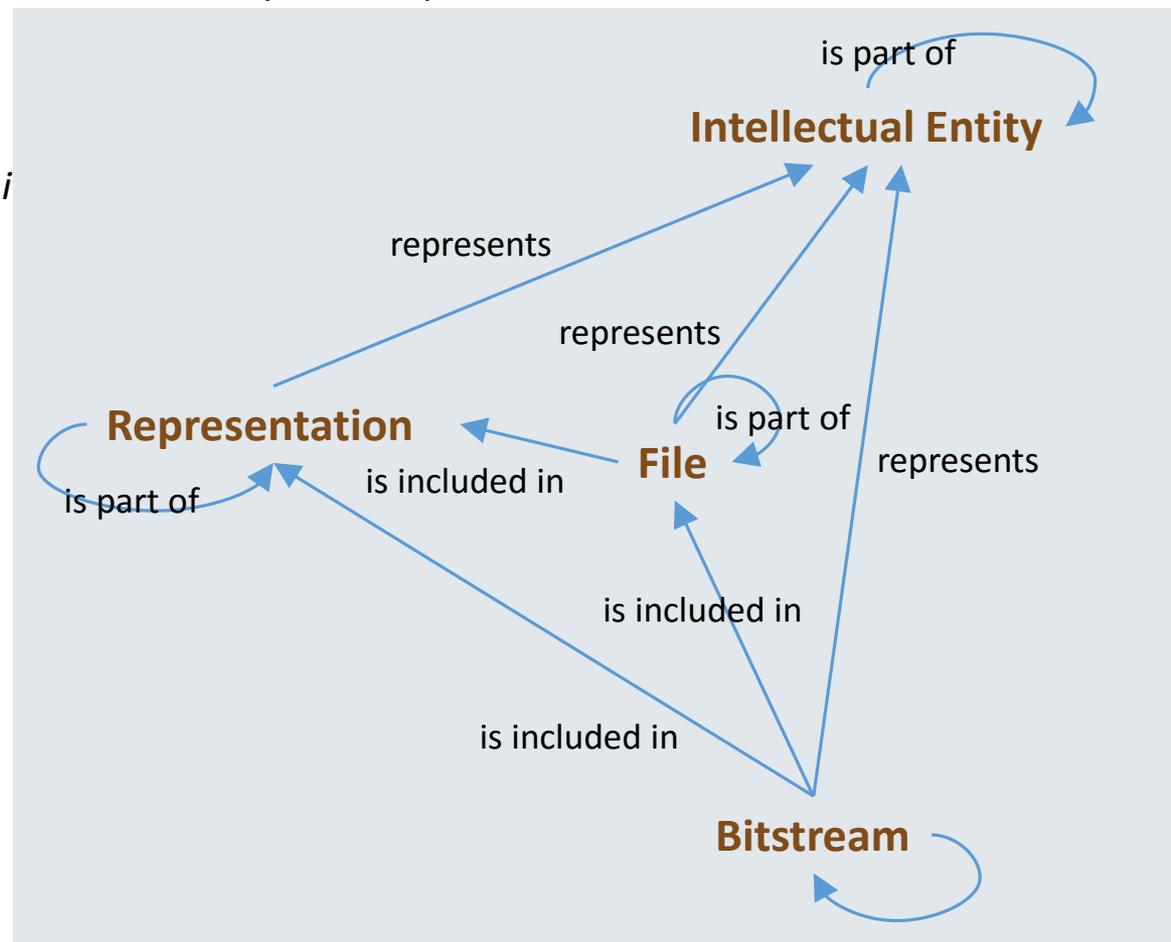
## Il modello dei dati Versione 3.0

**Entità:** sono “le cose” di importanza rilevante, dal punto di vista della conservazione e sono descritte dai metadati di conservazione - Intellectual Entity, File, Representation, Bitstream

- nel grafico costituiscono i nodi

**Relazioni tra Entità**

- nel grafico costituiscono gli archi



### Properties of Entities:

nel PREMIS DD

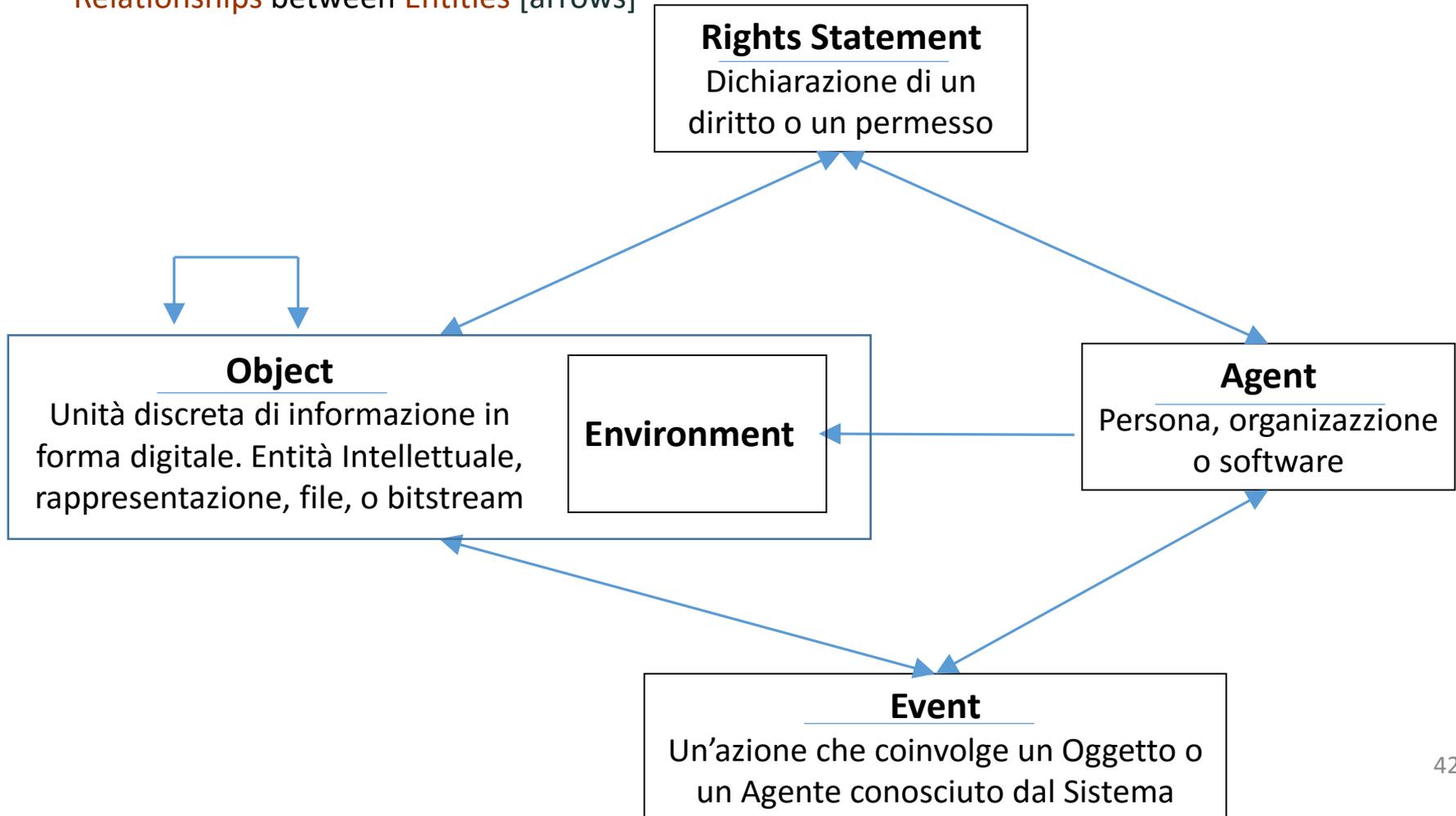
sono le unità semantiche che forniscono i dati necessari a descrivere ogni Entità

## Il modello dei dati Versione 3.0

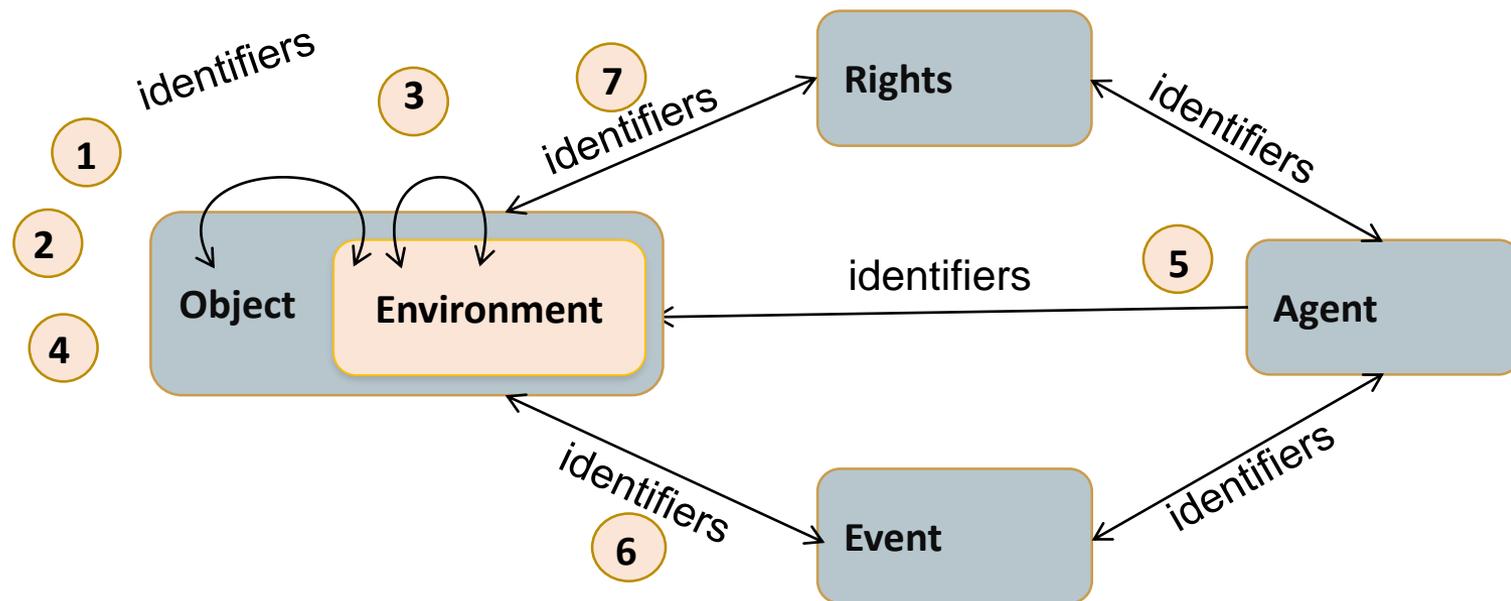
**Entities**, “things” relevant to digital preservation that are described by preservation metadata (Intellectual Entities, Objects, Events, Rights, Agents) [boxes]

**Properties of Entities** (semantic units)

**Relationships** between **Entities** [arrows]

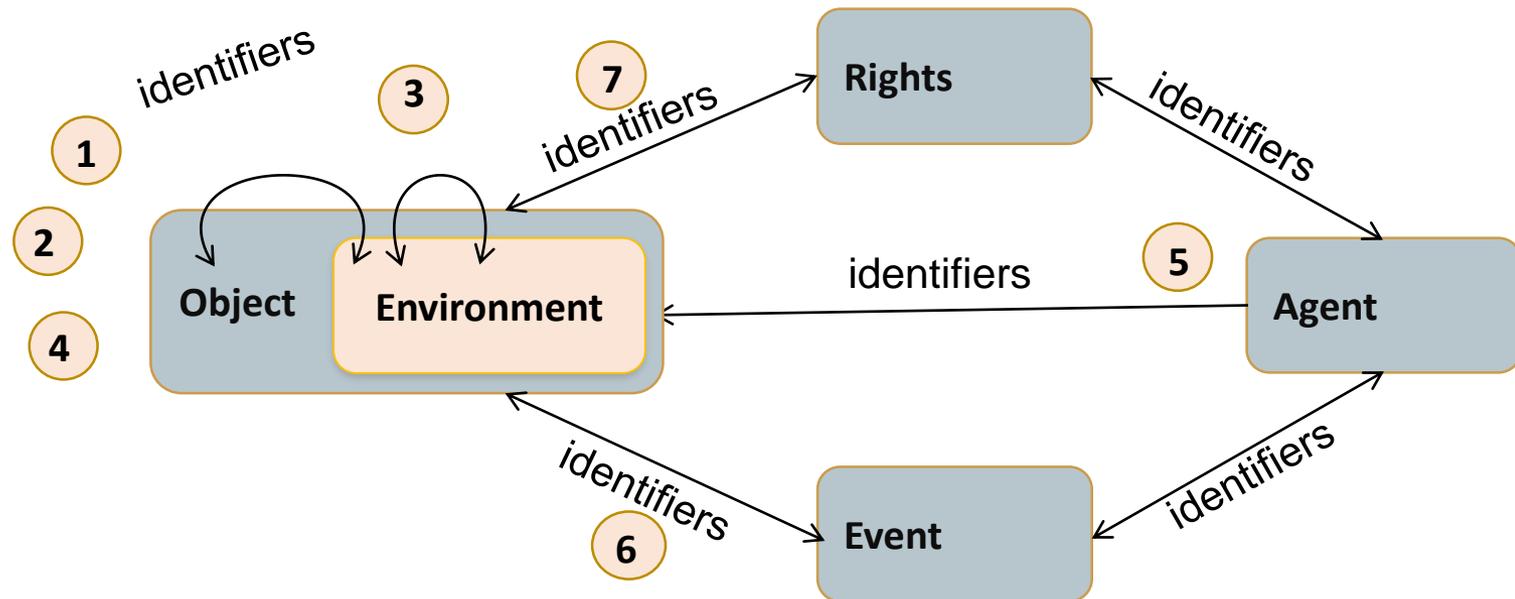


## Il modello dei dati Versione 3.0

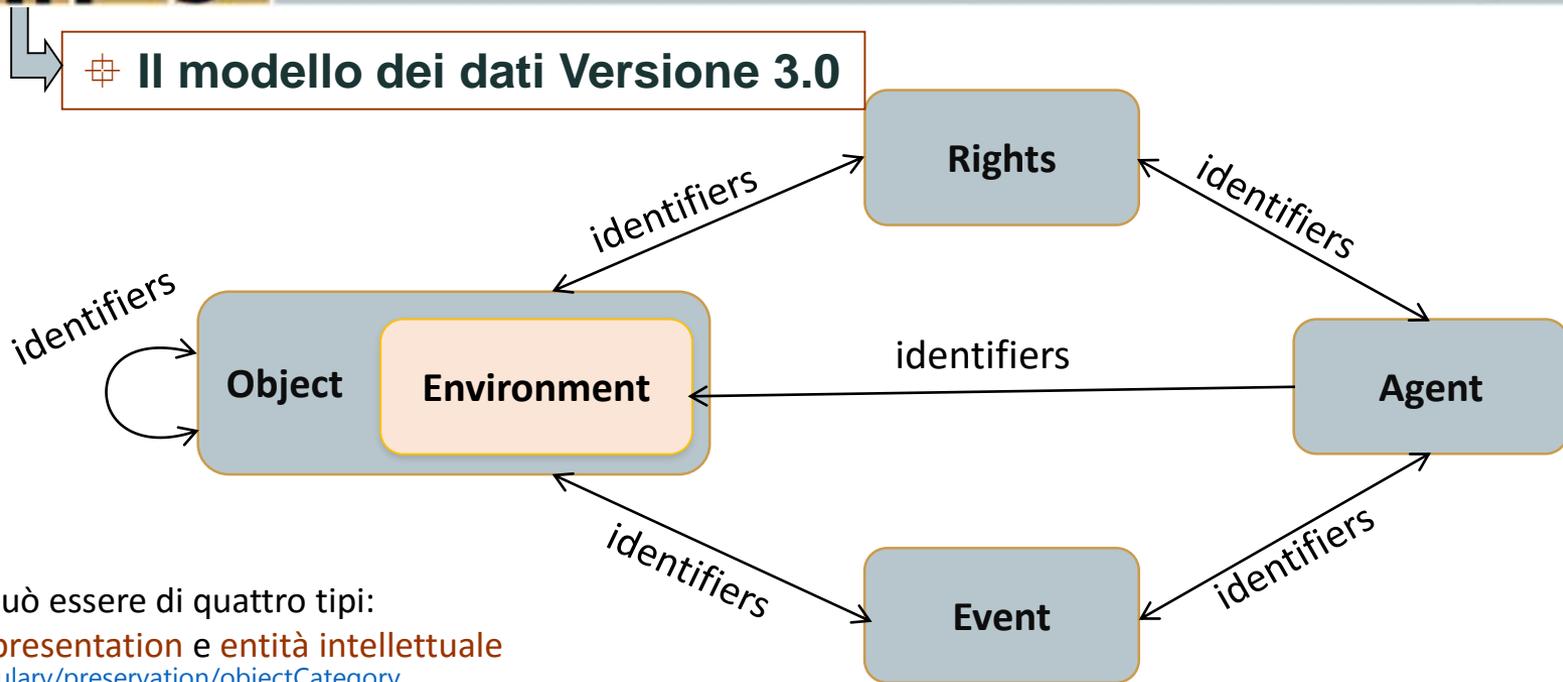


1. Un Oggetto (**Object**) è correlato ad un ambiente (environment) allo scopo di specificare il suo contesto computazionale (per esempio un **File** che è correlato all'applicazione software che lo "renderizza" (restituisce, traduce, rappresenta). L'ambiente nella versione 2 di PREMIS, viene implementato come un'unità semantica di tipo contenitore.
2. Un ambiente viene correlato ad un altro Oggetto (per esempio, una parte della pila di renderizzazione viene correlate alla documentazione di support all'uso, come ad esempio manuali utente o specifiche).
3. Un ambiente viene correlato ad un altro ambiente attraverso diversi tipi di relazione: inclusione, dipendenza, derivazione o altre relazioni (per esempio ambienti che interagiscono all'interno della pila di renderizzazione, come applicazioni, software di tipo plugins, sistemi operativi, hardware or software driver).

## Il modello dei dati Versione 3.0



4. Un ambiente è un Oggetto (per esempio, il codice sorgente del software) ed è preservato come entità di prima classe con sue caratteristiche proprie.
5. Un ambiente quando assume il ruolo di Agente (per esempio, un Agente software quando viene usato per una operazione di migrazione di formato come Evento di conservazione) viene descritto e conservato come IP (Pacchetto informativo OAIS).
4. Un ambiente ha un Evento associato al suo ciclo di vita (per esempio, la creazione dell'ambiente stesso, l'aggiunta di memoria RAM al computer o una versione del software).
5. Un ambiente ha una dichiarazione di diritti (RightsStatement) associata (per esempio, una licenza software o una policy).



L'entità Oggetto può essere di quattro tipi:  
**file**, **bitstream**, **representation** e **entità intellettuale**  
<http://id.loc.gov/vocabulary/preservation/objectCategory>

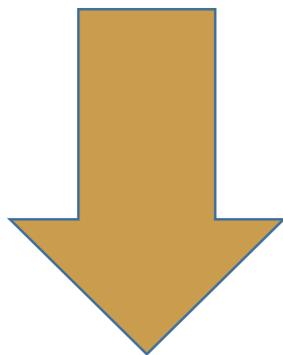
Un **intellectual entity** è un insieme coerente di contenuto che viene descritto come un'unità: ad esempio, un libro, una mappa, una fotografia, un seriale.

Un **file** è una sequenza di bytes ordinata e specificamente denominata che è conosciuta al sistema operativo. Un file potrebbe essere di 0 o più bytes ed ha un formato di file, permessi di accesso, e caratteristiche proprie del file system come le dimensioni (size) e la l'ultima data di modifica.

Un **bitstream** è costituito da dati contigui o non-contigui all'interno di un file che ha proprietà significative e comuni per gli scopi della conservazione. Non può essere trasformato in un file a se stante senza l'aggiunta di una struttura di file (headers, ecc.) e/o la sua riformattazione per essere conforme ad uno specifico formato di file.

Una **representation** è l'insieme dei file, inclusi i metadati strutturali, che sono necessari alla completa traduzione (rendition) di un'Entità Intellettuale

✦ Strumenti di supporto



**Il dizionario dei dati**

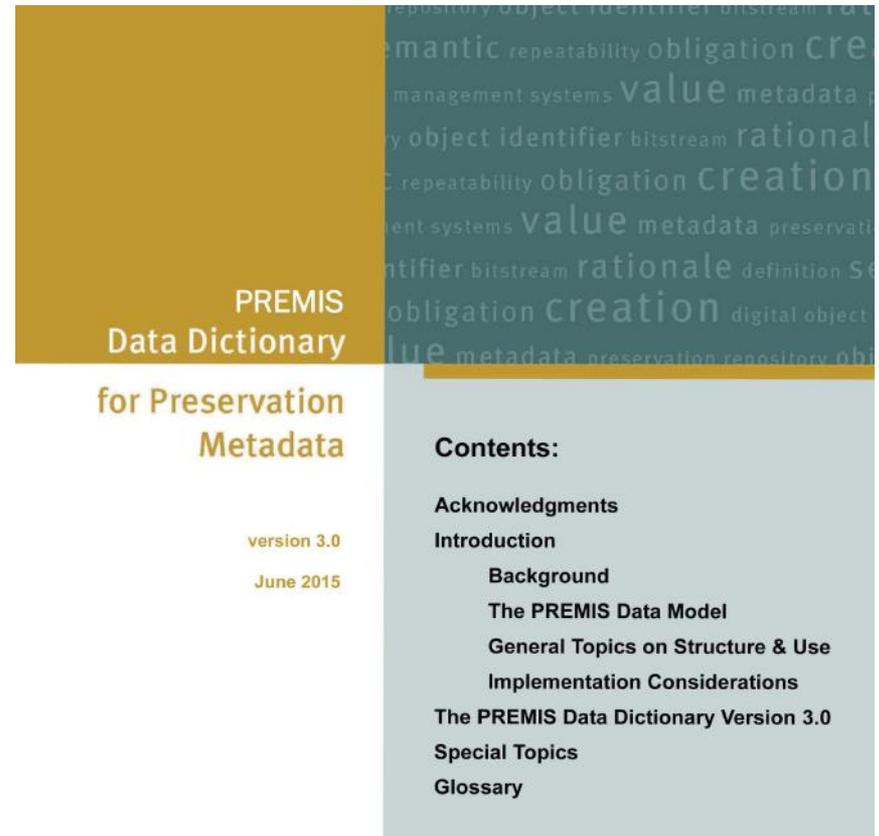
## Il Dizionario dei dati

**Maggio 2005:** *Data Dictionary for Preservation Metadata: Final Report of the PREMIS Working Group*

**Marzo 2008:** *PREMIS Data Dictionary for Preservation Metadata, version 2.0*

**Gennaio 2011:** *PREMIS Data Dictionary for Preservation Metadata, version 2.1*

**Giugno 2015:** *PREMIS Data Dictionary for Preservation Metadata, version 3.0*





## Il Dizionario dei dati

L'obiettivo primario del Dizionario dei Dati è stato definire un insieme di metadati di conservazione che:

- *Supporta la viabilità, renderabilità, comprensibilità, autenticità e identità degli oggetti digitali nel contesto della conservazione;*
- *Rappresenta l'informazione che la maggior parte dei depositi di conservazione ha bisogno di conoscere per preservare i materiali digitali nel lungo termine;*
- *Enfatizza l'aspetto dell'implementabilità dei metadati: definiti rigorosamente, supportati da linee guida per la creazione, la gestione e l'uso e orientato ai flussi di lavoro automatizzati; e*
- *Supporta la neutralità tecnica: nessun presupposto riguardo a tecnologie, strategie, immagazzinamento e gestione dei metadati ecc., viene preso in considerazione.*

Il Dizionario dei Dati PREMIS definisce “preservation metadata” come *l'informazione che un deposito usa per supportare il processo di conservazione digitale.*

*Data Dictionary for Preservation Metadata: PREMIS version 3.0 [PREMISDD-3.0]*

[PREMISDD-3.0] : Introduction



## Il Dizionario dei dati

### Il concetto di „Implementabilità“

- La maggior parte dei depositi gestisce larghe quantità di dati.
- Pertanto un fattore chiave per l' "implementabilità" dei metadata di conservazione è se il valore dei contenuti dei metadata possa essere automaticamente processato dal deposito.
- Laddove possibile il dizionario dei dati definisce unità semantiche per i metadata che non richiedono l'intervento umano per essere caricati o analizzati.
- Ad esempio i valori controllati per le lista di autorità sono preferite alle descrizioni testuali.



## Il Dizionario dei dati

### **“Implementable, core preservation metadata”**

*cose che la maggior parte dei depositi di conservazione deve gestire per supportare la conservazione digitale*

### **CORE**

- assolutamente richiesti in ogni circostanza – in qualsiasi fase del workflow devono essere richiesti;
- applicabili per qualsiasi tipo di deposito che implementi una qualsiasi strategia di conservazione;
- non necessariamente significa che essi sono obbligatori

### **IMPLEMENTABLE**

- definiti in modo rigoroso;
- supportati da linee guida e raccomandazioni;
- enfasi sui flussi di lavoro automatizzati (automaticamente prodotti e processati dal sistema di deposito);
- valori codificati da liste di autorità sono preferiti rispetto alle descrizioni testuali;
- neutralità tecnica multi-livello: nessun sistema, nessuna metodologia di gestione dei dati, nessuna strategia di conservazione viene presunta.

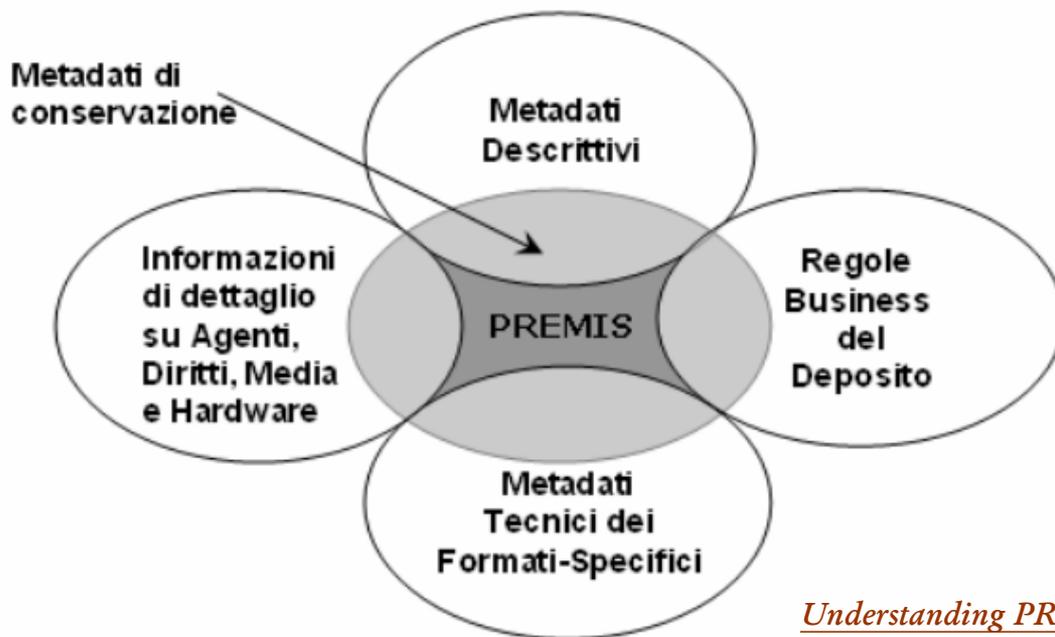
*Dal momento che l'enfasi è posta sul **bisogno di sapere** piuttosto che sul bisogno di memorizzare o rappresentare in un modo particolare, il WG del PREMIS ha preferito usare il termine “unità semantiche” (semantic unit) invece che “elemento di metadatozione” (metadata element).  
Il Dizionario dei Dati specifica la descrizione di unità semantiche.*

## Core preservation metadata

### “Nucleo implementabile di metadati di conservazione”

*cose che la maggior parte dei depositi di conservazione ha probabilmente bisogno di sapere per supportare i processi di conservazione digitale*

[PREMISDD-3.0] : p. 2



- Espande il numero di categorie tipicamente usate per differenziare i tipi di metadati: amministrativi (che includono rights e permessi), tecnici, e strutturali.
- Documenta per quanto possibile la provenienza digitale (la storia di un oggetto).
- Documenta le relazioni tra i diversi oggetti del deposito digitale.

*Understanding PREMIS* by Priscilla Caplan for the Library of Congress  
<http://www.loc.gov/standards/premis/understanding-premis.pdf> pp. 4-5

*Capire PREMIS* Traduzione italiana a cura di Angela Di Iorio  
[http://www.loc.gov/standards/premis/Understanding-PREMIS\\_italian.pdf](http://www.loc.gov/standards/premis/Understanding-PREMIS_italian.pdf)



## Il Dizionario dei dati

### Cosa è Dizionario dei Dati PREMIS (PREMIS-DD)

- Un modello comune di dati per organizzare/pensare metadati di conservazione
- Una guida per le implementazioni locali
- Uno Standard per scambiare pacchetti di informazione tra depositi
- Riferimento per il nucleo di metadati di conservazione

### Cosa NON è il PREMIS-DD

- Una soluzione completa: è necessario che i metadati di conservazione vengano popolati dal sistema di deposito
- Tutti i metadata necessari: esclude le regole di business, i metadati tecnici appartenenti a formati specifici, metadati descrittivi per l'accesso, metadati di conservazione "*non core*"
- Gestione del ciclo di vita degli oggetti fuori dal deposito
- Gestione dei diritti: si limita ai permessi riguardanti le azioni intraprese all'interno del deposito

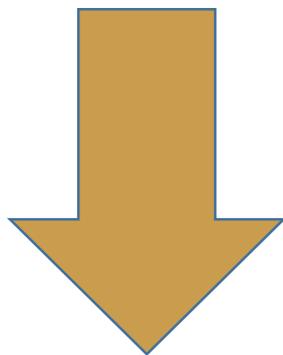


## Il Dizionario dei dati

### BENEFICI

- Fornisce una parte consistente dell'infrastruttura di conservazione digitale, e costituisce un blocco fondamentale sul quale poter costruire strategie di implementazione per la conservazione digitale efficace e sostenibile;
- costituisce la prima specifica tecnica omnicomprensiva per i metadati di conservazione prodotti da un processo costruito sul consenso a livello internazionale e interdisciplinare;
- è ampiamente applicabile in tutte le istituzioni, in tutti i contesti di conservazione digitale e in tutti i sistemi di implementazione;
- è orientato all'implementazione pratica;
- è supportato dalla PREMIS Maintenance Activity, che costituisce un riferimento centrale per informazioni, risorse per il PREMIS ed ospita la lista di discussione per il gruppo di implementatori del PREMIS.

## ✦ Strumenti di supporto



**Il sito Web**



## ⊕ L'attività di mantenimento

### Sito Web

- Presenza permanente nel Web, ospitato dalla Library of Congress
- Destinazione centrale per le informazioni relative al PREMIS, avvisi e altre risorse
- Informazioni riguardanti la lista di discussione del gruppo di implementazione di PREMIS [PREMIS Implementers' Group (PIG)]

### PREMIS Editorial Committee

- Definisce l'insieme di direttive/priorità per lo sviluppo del PREMIS
- Valuta le proposte di cambiamento e di sviluppo
- Coordina le revisioni del Dizionario dei Dati e dello schema XML

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



## Documentazione di supporto

**Sito web ufficiale:** <http://www.loc.gov/standards/premis/>

**PREMIS XML schema:** <http://www.loc.gov/standards/premis/premis.xsd>

Gruppo di discussione degli implementatori PREMIS: [pig@loc.gov](mailto:pig@loc.gov).

Per iscriversi, mandare un'e-mail all'indirizzo [listserv@loc.gov](mailto:listserv@loc.gov)  
con oggetto: "subscribe pig [il tuo nome]"

Per eventuali commenti e questioni a [premis@loc.gov](mailto:premis@loc.gov).

**Vocabolari controllati per termini della conservazione digitale:**

<http://id.loc.gov/preservationdescriptions/>



## Bibliografia

**OAIS:** <http://public.ccsds.org/publications/archive/650x0m2.pdf>

**Lettura pre-corso:** [http://www.loc.gov/standards/premis/Understanding-PREMIS\\_italian.pdf](http://www.loc.gov/standards/premis/Understanding-PREMIS_italian.pdf)

### Risorse informative a supporto di PREMIS:

PREMIS homepage: <http://www.loc.gov/standards/premis/index.html>

PREMIS versione 2 homepage: <http://www.loc.gov/standards/premis/v2/index.html>

PREMIS versione 3 homepage: <http://www.loc.gov/standards/premis/v3/index.html>

Ontologia PREMIS 2.2 OWL - documentazione: <http://www.loc.gov/standards/premis/ontology-announcement.html>

Ontologia PREMIS 2.2 OWL - namespace loc.gov: <http://id.loc.gov/ontologies/premis.html#>

Conformità al PREMIS Conformance - 2010:

<http://www.loc.gov/standards/premis/premis-conformance-201010.pdf>

Conformità al PREMIS - 2015: <http://www.loc.gov/standards/premis/premis-conformance-20150429.pdf>

<http://www.loc.gov/standards/premis/premis-conformance-20150429.pdf>

Using PREMIS with METS:

<http://www.loc.gov/standards/premis/premis-mets.html>

**PREMIS XML schema** PREMIS XML-schema versione corrente (3.0): <http://www.loc.gov/standards/premis/premis.xsd>

PREMIS XML-schema versione 2.3: <http://www.loc.gov/standards/premis/v2/premis-v2-3.xsd>

PREMIS XML-schema versione 2.2: <http://www.loc.gov/standards/premis/v2/premis-v2-2.xsd>

**Esempi:** <http://www.loc.gov/standards/premis/examples.html>

### Vocabolari controllati:

Library of Congress preservation vocabularies: <http://id.loc.gov/vocabulary/preservation>

### Datetime:

Extended datetime format: <http://www.loc.gov/standards/datetime/>

### Extension schema:

Example rights schema: [www.cdlib.org/inside/projects/rights/schema/](http://www.cdlib.org/inside/projects/rights/schema/)

### Mix:

Mix homepage: <http://www.loc.gov/standards/mix/>

### METS:

METS homepage: <http://www.loc.gov/standards/mets/>

**Altre fonti di riferimento:** <http://www.oclc.org/content/dam/research/activities/trustedrep/repositories.pdf>