

# Um modelo de dados para metadados de preservação

Em Junho de 2003 OCLC e RLG criaram um grupo de trabalho internacional chamado PREMIS (PReservation Metadata Implementation Strategies) que funcionou durante dois anos. Este grupo de trabalho incluía mais de trinta especialistas de cinco países diferentes, representantes das bibliotecas, museus, arquivos, Governos e sector privado. O objectivo era definir «os grupos “essenciais” de metadados de preservação» utilizados pela comunidade de preservação digital. Em Maio 2005, foi publicado um relatório final. Incluía um modelo de dados para metadados de preservação e um dicionário de dados versão 1.0. Actualmente, a actividade de implementação é assegurada por um grupo de manutenção PREMIS que está encarregue da revisão e da manutenção do modelo e do dicionário de dados.

## As necessidades

O grupo de trabalho identificou cinco domínios pertinentes para os metadados de preservação:

- A proveniência: este domínio deverá conservar as informações de qualidades históricas da conservação do objecto digital, depois da sua criação e as alterações seguintes da conservação física e/ou de propriedade.
- A autenticidade: este domínio deverá englobar as informações necessárias para atestar que o objecto digital é o que afirma ser, que não foi alterado, seja intencionalmente, seja por inadvertência, sem documentar o facto.
- As actividades de preservação: este domínio deverá compreender a documentação das acções sucessivas transacções para a preservação do objecto digital, e guardar as pistas das consequências destas acções na apresentação, no fundo, o resumo da funcionalidade deste objecto.
- O ambiente técnico: deverá descrever o material, o sistema de exploração e os softwares necessários à activação e à utilização do objecto no estado em que se encontra conservado.
- A gestão dos direitos: devem ser conservados todos os direitos restritivos da propriedade intelectual que limitem o poder do depósito de conservação de tomadas de decisão sobre as disposições para preservação do objecto digital e a sua disseminação perante os utilizadores presentes e futuros.

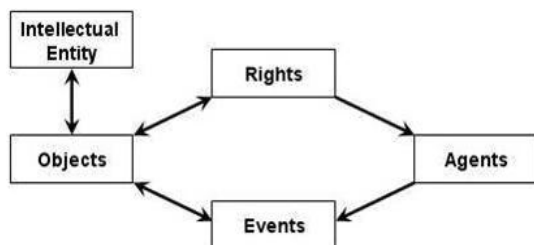
## Resumo do modelo de dados

Aqui resume-se o modelo de dados.

O modelo de dados PREMIS é constituído por entidades, por relações e por propriedades que são denominadas por unidades semânticas.

### As entidades

- A entidade intelectual– um conjunto coerente de conteúdos que são razoavelmente descritos como uma unidade, por exemplo um livro, um mapa, uma fotografia ou uma base de dados. Porque uma entidade está bem descrita nos metadados descritivos, é considerada fora de campo do dicionário de dados.
- Objecto ou objecto digital– uma unidade discreta de informação na sua forma digital.
- Evento- uma acção que implica no mínimo um objecto ou um agente conhecido do depósito de preservação.



- Agente– uma pessoa, um organismo ou um software associado aos eventos de preservação durante o tempo de vida do objecto.
- Os direitos– as asserções ligadas a um ou vários direitos ou permissões pertencentes a um objecto e/ou a um agente.

### As relações

As relações são fórmulas de associação das ocorrências entre as entidades. «Relação» pode ser interpretada, num sentido lato ou restrito e qualquer relacionamento pode ser formulado de várias maneiras diferentes. As relações entre objectos têm três variações de base:

- As relações de estrutura demonstram as relações entre os objectos e as suas partes. As relações de estrutura entre os ficheiros que constituem uma representação de uma entidade intelectual são, claramente de metadados de preservação essenciais. Se num depósito de preservação as partes de um objecto digital não puderem ser recuperadas em conjunto, então este objecto não está preservado.

## Informações e Fontes complementares

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/>

- As relações de derivação resultam da replicação ou da transformação de um objecto. O conteúdo intelectual de um objecto resultante é o mesmo, mas a instanciação do objecto e do seu formato são, sem dúvida, diferentes. Vários objectos digitais são complexos e as informações estruturais e de derivação podem mudar no tempo, bem como as actividades de preservação.
- Quando existe uma relação de dependência, impõe um objecto a outro para reforçar a sua função, os resultados ou a coerência do seu conteúdo. O objecto de complemento ou de ajuda não faz formalmente parte do objecto em si mas é necessária a sua apresentação ou a sua execução.

### As propriedades

As unidades semânticas são as propriedades de uma entidade. Em certos casos, uma unidade semântica pode ser um receptáculo que reagrupa um conjunto de unidades semânticas da mesma família; as subunidades agrupadas são agora denominadas de componentes semânticas da unidade semântica.

### O princípio de 1:1

The Para os metadados, o princípio 1:1 que coloca cada descrição não descrita, é uma e apenas um activo. Aplicado aos metadados de PREMIS, cada objecto inserido num depósito de preservação (ficheiro, cadeia de bits, representação) está agora descrito como uma cadeia de bites estatísticos. Não é possível alterar este conjunto mas apenas criar um novo que esteja conectado a um objecto fonte por uma relação de derivação. Com efeito, o dicionário de dados é uma unidade semântica unicamente pela data de modificação de um objecto desde que, por definição, um objecto não possa ser modificado.

### Do modelo ao dicionário

Do modelo de dados a um dicionário de dados associados que compreende todas as unidades semânticas pertinentes descreve quatro entidades cobertas pelo dicionário (objectos, agentes, eventos, direitos). No modelo de dados, as relações entre entidades de tipos diferentes são materializadas por setas.

O dicionário de dados trata-os como informação de ligação, incluindo a informação da entidade A relacionando-a com a entidade B. No modelo de dados, cada entidade tem um identificador único que pode ser utilizado como indicador. Por exemplo, a entidade objecto tem setas que apontam para as entidades Intelectual e Acontecimentos. São apresentados no dicionário de dados pelas unidades semânticas «Identificador de ligação da entidade intelectual» e «identificador da entidade Acontecimentos».

### As vantagens e os inconvenientes

O dicionário de dados PREMIS é o resultado de um processo de construção consensual, inter-domínios e internacional; o que aumenta suas hipóteses de ser amplamente utilizável por todo o tipo de instituições, de contextos de preservação digital e de implementações em diferentes sistemas informáticos. Como um padrão de metadados de preservação, deverá poder cobrir a totalidade das necessidades actuais. Está orientado para implementações concretas e para a interoperabilidade, para facilitar as manipulações de objectos e cumprir a sua função primordial: documentar os objectos digitais através dos tempos e torná-los acessíveis a longo prazo. O inconveniente do dicionário de dados PREMIS é que ninguém poderá dizer com certeza que eficácia terá. O impacto dos desenvolvimentos futuros na utilização dos objectos é difícil de prever.

### Conclusões

O dicionário de dados é uma peça crucial da infra-estrutura da preservação digital; também será uma componente com a qual se podem implementar estratégias duradouras e eficazes de preservação digital.

**Author:** Angela Di Iorio, *Fondazione Rinascimento Digitale*

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

**Translation:** Sofia Pina, [sofia.opina@gmail.com](mailto:sofia.opina@gmail.com)