

**Francesca Tomasi**

*La preservazione del contenuto degli oggetti  
culturali: formalizzare la provenance*

*Dalla preservazione alla curation*

È dallo storico report della *Task Force on Archiving of Digital Information* (Garret - Waters 1996) che il tema della preservazione delle risorse digitali, o anche degli oggetti culturali (digitali), è al centro del dibattito.<sup>1</sup> Perché siamo di fronte ad un concetto che travalica i confini della mera conservazione fisica, per abbracciare il più ampio principio della valorizzazione della memoria culturale (Asmann 1999 [2002]). Preservare il ricordo, collettivo e individuale, attraverso l'accesso, garantito nel tempo, al patrimonio è il ruolo che le istituzioni culturali sono chiamate a ricoprire. E se la trasmissione del sapere avviene attraverso processi di dematerializzazione della conoscenza, che richiedono al digitale di svolgere in prima istanza un ruolo

---

<sup>1</sup> La bibliografia sul tema è davvero sconfinata. Un ottimo punto di partenza è Corrado - Sandy 2004, o ancora, in particolare in prospettiva testo-centrica, Deegan - Tanner 2006. Sul fronte dei progetti istituzionali merita un particolare rilievo l'handbook *DCH-RP* 2013, cui partecipa l'ICCU, all'interno del progetto Digital Meets Culture. Per un contributo più recente, si veda il volume online della Digital Preservation Coalition 2015.

lo di surrogato dell'artefatto culturale, la creazione di oggetti digitali diventa un tema di larga portata. Perché preservare oggetti digitali significa preservare la memoria. E la preservazione è un'attività che non riguarda solamente la gestione del problema dell'obsolescenza degli apparati (hardware) e degli strumenti di accesso al sapere (software), ma è un processo che parte ben prima dell'obsolescenza, ovvero nasce con la nascita dell'oggetto digitale, assieme agli elementi, o anche descrittori, necessari alla sua qualificazione e identificazione (cfr. Tomasi 2010). Ecco allora che preservare il digitale non riguarda solo la messa in atto di processi conservativi, ma inizia dal concepimento, e dalla creazione (che implica la selezione di formati, vocabolari e linguaggi formali). La scelta delle strategie di rappresentazione del sapere ha un impatto profondo sulle possibilità della sua preservazione.

Non dobbiamo però dimenticare che l'oggetto culturale digitale è un'entità informativa complessa, ovvero anche multilivellare: ogni livello è una sfaccettatura dello stesso oggetto, che contribuisce all'integrità dell'entità/unità informativa. Ecco che la preservazione se da un lato deve saper prevedere e valutare le più opportune strategie conservative per ogni livello, allo stesso tempo deve considerare l'esigenza di garantire l'accesso continuativo nel tempo all'oggetto come insieme di livelli dialoganti. Se condividiamo la posizione di Webb (2003), riconosciamo che l'oggetto digitale va analizzato secondo quattro punti di vista, ovvero anche quattro livelli:

- come oggetto fisico, sequenze di bit caricati su un certo *medium* (problema del deterioramento dei supporti);
- come oggetto logico, un codice leggibile dalla macchina la cui esistenza non è legata a un particolare supporto fisico di conservazione (problema dell'obsolescenza delle tecnologie hardware e software);
- come oggetto concettuale, dotato di un contenuto e quindi di un significato, comprensibile da parte di un lettore umano;
- come una collezione, o anche insieme, di elementi eterogeni che contribuiscono a determinare il contenuto dell'oggetto digitale.

Possiamo anche dire che l'oggetto digitale è concepito come un'entità a tre livelli: fisico (flusso di bit), logico (i formati) e semantico (il contenuto, comprese le informazioni contestuali), chiamato anche concettuale (Thibodeau 2002). Questi tre livelli dialoganti e interconnessi contribuiscono a qualificare la nozione di 'collezione', che si esplica nel rapporto dell'oggetto con gli altri oggetti/elementi utili a rafforzarne lo spessore semantico. E la preservazione di conseguenza è una suite di azioni che devono contemplare la multilivellarità dell'oggetto e le relazioni veicolate.

La conservazione delle risorse digitali riguarda allora la messa in atto di attività e di procedure, nonché la riflessione sugli strumenti e sui metodi, finalizzati ad assicurare che gli oggetti digitali siano accessibili, utilizzabili e identificabili nel medio e lungo periodo.

Questo significa anche modellizzare ad un alto livello di astrazione. Secondo le note specifiche del *reference model* di OAIS (2012), la preservazione è un problema di definizione di un *framework* che detti un modello di gestione degli oggetti digitali, secondo pacchetti informativi diversificati ("submission information package", "archival information package", "dissemination information package") e entità funzionali, ovvero azioni ("ingest", "storage", "data management", "administration", "preservation planning" e "access") finalizzate alla *long-term preservation*.

Ma è in particolare sulla dimensione concettuale, ovvero anche sul contenuto, che è necessario riflettere, perché l'aspetto determinante del principio di preservazione del digitale è quello, come si è già detto, della conservazione del sapere, ovvero di tutela della memoria. Non è un caso se negli ultimi tempi l'attenzione delle istituzioni deputate alla conservazione del patrimonio, anche digitale, si è spostato dal tema della preservazione a quello della *digital curation*, vale a dire della 'cura' dell'intero ciclo vitale degli oggetti digitali, che significa la messa in atto non di singole azioni di tutela, ma di una serie combinata di attività congiunte e combinate che si esplicano nella forma di un processo: «Nel mese di Agosto 2011 l'ALA (American Library

Association) ha aperto la mailing list *Digital Curation Interest Group*,<sup>2</sup> che oggi conta più di 500 iscritti» (Testoni 2013). Ed è dall'ottima e condivisa definizione del DCC (Digital Curation Center) che dobbiamo partire: «Digital curation involves maintaining, preserving and adding value to digital research data throughout its lifecycle».<sup>3</sup> Ovvero la digital curation riguarda sì il mantenimento e la preservazione, ma richiama anche l'esigenza di valorizzare gli oggetti, o anche i *digital research data*, lungo tutto il loro ciclo di vita. Tale ciclo di vita, secondo il modello elaborato dal DCC,<sup>4</sup> può essere riassunto in:<sup>5</sup>

- concettualizzazione: la creazione di dati digitali va prima pianificata, includendo nella pianificazione i metodi di cattura e le opzioni di immagazzinamento (*storage*);

- creazione o ricezione: è la fase in cui il dato viene prodotto, o ricevuto 'grezzo' dall'autore e comporta l'assegnazione dei diversi metadati (si noti: senza l'assegnazione di metadati non c'è creazione di dati);

- valutazione e selezione: è necessario selezionare gli oggetti digitali che necessiteranno di una conservazione a lungo termine, in ottemperanza a linee guida, regole e requisiti legali;

- acquisizione (*ingest*): è il momento in cui si trasferiscono gli oggetti digitali in un archivio o in un *repository*;

- azione di conservazione: si tratta di assicurare la conservazione a lungo termine e la preservazione dell'integrità (*retention of the authoritative nature*) dei dati;

- *store*: i dati devono essere conservati in modo sicuro e aderente agli standard;

---

<sup>2</sup> Digital Curation Interest Group, <<http://lists.ala.org/sympa/info/acr-igdc-l>>.

<sup>3</sup> Sul DCC (<<http://www.dcc.ac.uk/digital-curation/what-digital-curation>>) si veda in particolare Higgins 2011.

<sup>4</sup> DCC Curation Lifecycle Model, <<http://www.dcc.ac.uk/resources/curation-lifecycle-model>>.

<sup>5</sup> Trad. del modello tratta da Testoni 2013.

- accesso e uso e riutilizzo: è necessario assicurarsi che i dati siano accessibili per l'uso quotidiano da parte degli utenti designati;
- trasformazione: la creazione di nuovi dati dagli originali, sia attraverso una rielaborazione (ad esempio creazione di subset) sia, eventualmente, attraverso migrazione in un differente formato.

Ecco dunque alcuni concetti fondamentali: il tema della concettualizzazione, qui espressa nei termini di una pianificazione del momento concettuale, è l'atto che precede la creazione; l'importanza dei metadati nel processo di realizzazione della risorsa; la preservazione intesa come uno degli aspetti del processo; il senso della valorizzazione del principio di integrità, ovvero anche dell'*authoritativeness*. Nelle "preservation actions" previste dal DCC leggiamo che esse: «should ensure that data remains authentic, reliable and usable while maintaining its integrity»; ovvero anche che:

digital preservation is about more than keeping the bits – those streams of 1s and 0s that we use to represent information. It is about maintaining the semantic meaning of the digital object and its content, about maintaining its provenance and authenticity, about retaining its 'interrelatedness', and about securing information about the context of its creation and use (Ross 2012, p. 44).

Desumiamo allora un importante principio dal quale vogliamo partire: la dichiarazione di *provenance* è uno degli strumenti finalizzati a garantire la preservazione non solo dell'oggetto digitale come sequenza logica binaria, ma del suo essere dotato di un contenuto intellegibile. E a garantire autenticità, integrità, affidabilità<sup>6</sup> degli oggetti

---

<sup>6</sup> Sul tema si veda in particolare la serie dei documenti prodotti in seno al progetto InterPARES - International Research on Permanent Authentic Records in Electronic Systems (<<http://www.interpares.org/>>), avviato nel 1998 e giunto ad oggi alla sua fase 4 (InterPARES Trust, <<https://interparestrust.org/>>). Un'altra utile risorsa informativa è il recente Centro Italiano sulla Conservazione digitale (<<http://www.conservazionedigitale.org/>>, 2014-), in particolare la sezione rela-

digitali, ovvero anche dei dati, concorrono i metadati, in particolare i metadati sulla *provenance*. Affinché ci sia preservazione è allora a partire dal momento della creazione della risorsa digitale, ma poi lungo tutto il suo ciclo di vita, che devono essere previsti meccanismi atti a garantire la tutela dell'oggetto culturale attraverso le informazioni necessarie alla sua conservazione e in vista della sua valorizzazione.

### *Provenance e metadati*

Ecco dunque che una delle soluzioni adottate per assicurare la preservazione attraverso meccanismi di metadattazione è rappresentata dalla documentazione relativa alla *provenance*. E fra gli standard di metadati per la preservazione, PREMIS (2008) rappresenta senz'altro l'esperimento più compiuto. È nelle specifiche di PREMIS che gli autori dichiarano:

The provenance of the object: information describing the custodial history of the object, potentially stretching back to the object's creation, and moving forward through successive changes in physical custody and/or ownership. Provenance information includes descriptions of the actions that have been taken to preserve the object over time. Such information describes aspects of the digital preservation process used to maintain the object; it would also record any consequences of this process that alter the content, or look, feel, and functionality of the object. Related to this would be information that serves to establish and validate the object's authenticity, i.e., that the preserved object is in fact what it purports to be, and has not been altered, intentionally or unintentionally, in an undocumented way. Authenticity would include such elements as fixity and integrity (Lavoie - Gartner 2013, p. 5).

Ecco allora che se la *provenance* racconta l'origine di un oggetto digitale, allo stesso tempo essa dovrà garantire la documentazione di

---

tiva all'autenticità (<<http://www.conservazionedigitale.org/wp/approfondimenti/autenticita-e-validazione/>>).

ogni passo della catena evolutiva dell'oggetto, di ogni intervento di revisione, di ogni azione compiuta o anche di ogni evento accaduto che abbia coinvolto l'informazione veicolata da quell'oggetto.

Fino dalla concettualizzazione, documentare la *provenance* è garanzia di valorizzazione dell'oggetto culturale come entità complessa. La storia dell'oggetto, ovvero anche il contesto necessario a darne consistenza semantica, rappresenta il primo passo verso la garanzia di autorevolezza della fonte. Come ci insegna la diplomatica, solo lo studio critico del documento può assicurare il valore di testimonianza storica a quel documento, garantendone l'autenticità (Pratesi 1979). Analogamente, tanto in archivistica quanto in informatica – in qualità di discipline che si pongono criticamente nei confronti del concetto di *provenance* – documentare la provenienza significa garantire al fruitore di potersi fidare di un'informazione, di poter riusare quella informazione proprio perché è affidabile, e di poter dare credito a chi quell'informazione ha deciso di trasmettere.

Il modello dei dati di PREMIS<sup>7</sup> è particolarmente efficace per sintetizzare i concetti che governano la filiera della *provenance*: esiste un'entità intellettuale, o anche concettuale, che rappresenta il massimo livello di astrazione; esiste un oggetto, che è ciò che viene effettivamente realizzato e conservato; esistono degli eventi, che riguardano le azioni compiute su quell'oggetto ed esistono degli agenti, intesi tan-

---

<sup>7</sup> Cfr. PREMIS (Preservation Metadata Maintenance Activity), <<https://www.loc.gov/standards/premis/>>: «The PREMIS Data Dictionary for Preservation Metadata is the international standard for metadata to support the preservation of digital objects and ensure their long-term usability. Developed by an international team of experts, PREMIS is implemented in digital preservation projects around the world, and support for PREMIS is incorporated into a number of commercial and open-source digital preservation tools and systems. The PREMIS Editorial Committee coordinates revisions and implementation of the standard, which consists of the Data Dictionary, an XML schema, and supporting documentation». L'ultima versione del Data Dictionary, Version 3.0 è del 2015 (<<https://www.loc.gov/standards/premis/v3/premis-3-0-final.pdf>>).

to come persone, organismi e istituzioni quanto software, che hanno ricoperto un certo ruolo in un dato evento (cfr. fig. 1).

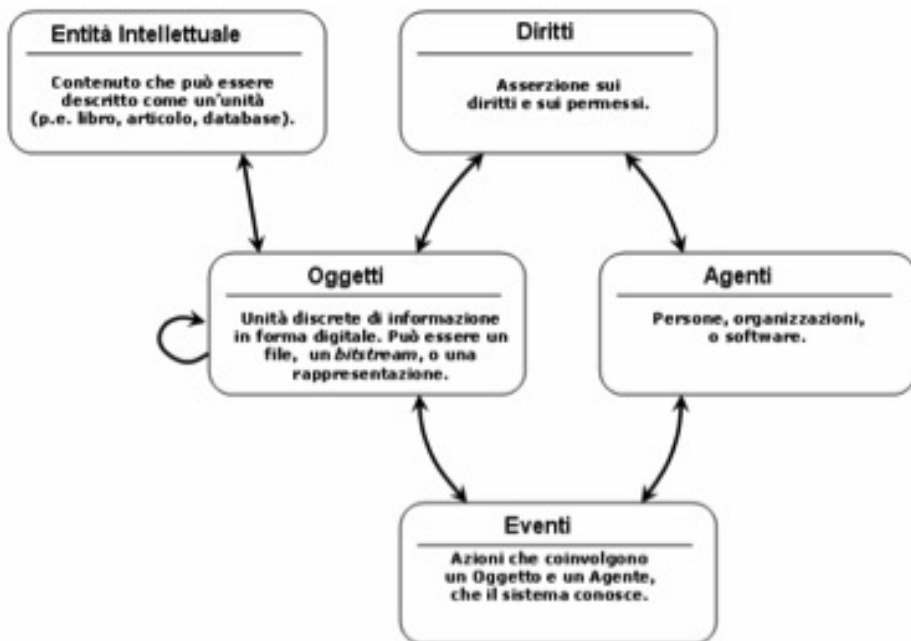


Fig.1. Modello dati Premis

Trad. it: *Capire PREMIS*, 2009

<[http://www.loc.gov/standards/premis/Understanding-PREMIS\\_italian.pdf](http://www.loc.gov/standards/premis/Understanding-PREMIS_italian.pdf)>

Ragionare in questi termini nel panorama del Word Wide Web, significa riconoscere che ogni documento, o meglio ogni dato, è un'entità semantica astratta, e che è necessario chiarire chi, e in quali circostanze, ha prodotto quel dato o quel documento. Nel mare magnum di informazione che il Web ogni giorno alimenta, i documenti e i dati, ovvero anche gli oggetti del patrimonio culturale, soffrono di due importanti problemi:



- l'accesso continuato nel tempo al contenuto informativo veicolato;
- la qualità dell'informazione e il grado di affidabilità del sapere trasmesso.

Il Web semantico ci sta abituando ad un nuovo approccio al sapere. In particolare con il principio dei Linked Open Data (LOD) si sta assistendo al passaggio da una realtà documento-centrica ad una dato-centrica. Questo significa che ai documenti si stanno sostituendo i dati, ovvero anche elementi atomici, identificati univocamente e che fanno riferimento a cose, nei termini di entità anche astratte (p.e. persone, luoghi, periodi ed opere). Ecco che ogni dato, che potenzialmente descrive una porzione di un oggetto culturale complesso, è un possibile livello della rappresentazione del sapere che quell'oggetto veicola. Dati e metadati si fondono nei LOD in un sistema che vede, parlando in termini RDF, l'entità identificata univocamente (il dato soggetto) associata ad un'altra entità (il dato oggetto) attraverso una relazione (il dato predicato). E l'oggetto culturale è un insieme di dati interconnessi che ne garantiscono il contenuto, ovvero ne specificano il livello semantico.

Ma la progressiva amplificazione della quantità dei dati, che vengono periodicamente prodotti, richiede un controllo sulla qualità dell'informazione. Mentre il LOD cloud aumenta esponenzialmente (Abele *et al.* 2017), vanno previsti meccanismi di preservazione del contenuto informativo, al fine di garantire l'accesso alla conoscenza *ad perpetuam memoriam*. La *provenance* diventa in questo contesto la formalizzazione di ogni azione compiuta sull'oggetto culturale nei termini di entità multilivellare. Se attraverso l'identificazione univoca, resa possibile dal meccanismo degli URIs, si vuole garantire l'accesso nel tempo alle risorse, attraverso la *provenance* si provvederà a garantire la qualità dell'informazione che quella risorsa intende trasmettere.

Ecco allora che in senso estensivo la *provenance* andrà concepita come qualunque atto formale di attribuzione di responsabilità di un'azione compiuta su di un oggetto o meglio sul suo contenuto. Sarà quindi lecito considerare anche il procedimento di creazione di una

tripla RDF come un atto a cui assegnare un *creator*. Attraverso l'attribuzione di paternità agli asserti è possibile riconoscere il responsabile della creazione del dato, e quindi anche della formalizzazione attraverso l'adozione di un modello di *resource description*. Questo approccio è qualificabile come un meta-livello che lega l'interpretazione del contenuto di una fonte all'autore, ovvero l'agente, dell'interpretazione stessa. Perché ogni dichiarazione finalizzata a costruire una relazione fra URIs, o fra un URI e un letterale, è un procedimento ermeneutico. Ogni dichiarazione, a sua volta, si basa sulla consultazione di fonti che consentono all'interprete di supportare il proprio asserto. Il contesto, inteso archivisticamente come il legame fra la fonte e il soggetto produttore o detentore di quella fonte, diventa allora il meccanismo per preservare il dato stabilendone la provenienza, avvicinando quindi il problema della conservazione logica a quello dell'autenticità. In senso estensivo il contesto lega l'oggetto a chi, su quell'oggetto, ha espresso un'idea, formulato un'opinione o veicolato un parere utile a esplicitare il contenuto dell'oggetto stesso, sulla base dell'assunzione di un individuale punto di vista o alla luce delle proprie conoscenze, ovvero della consultazione delle fonti (cfr. Daquino - Tomasi 2015).

Ma per un'autentica formalizzazione del principio di preservazione, è necessario passare dai metadati alle ontologie, ovvero anche alla concettualizzazione del principio di provenienza. Se la preservazione del contenuto è preservazione dell'informazione veicolata, la creazione di ontologie, ovvero anche la concettualizzazione, è preservazione della conoscenza.

### *Il ruolo delle ontologie nella formalizzazione della provenance*

Il ruolo delle ontologie nel contesto del Web semantico è cosa nota: è con le ontologie che viene esplicitato il significato, ovvero che vengono messi in atto meccanismi finalizzati alla comprensione da parte delle macchine. Anche lo studio della *provenance* è stato oggetto di riflessione ontologica, nei termini dell'esigenza di creare un modello

riusabile in contesti diversi. Formalizzando la *provenance* si preservano infatti concettualmente gli oggetti culturali, contribuendo a dar loro il valore di fonti attendibili.

Come nelle scienze del documento così in informatica:

*Provenance* is defined as a record that describes the people, institutions, entities, and activities involved in producing, influencing, or delivering a piece of data or a thing. In particular, the provenance of information is crucial in deciding whether information is to be trusted, how it should be integrated with other diverse information sources, and how to give credit to its originators when reusing it. In an open and inclusive environment such as the Web, where users find information that is often contradictory or questionable, provenance can help those users to make trust judgements.<sup>8</sup>

Così troviamo scritto nel *data model* di PROV-O,<sup>9</sup> l'ontologia ideata esattamente per rispondere all'esigenza di concettualizzare la *provenance*, ovvero di garantire la formalizzazione di quelle informazioni su entità ("entity"), attività ("activity") e agenti ("agent"), coinvolte nella produzione di dati, o cose; informazioni che possono essere utilizzate per stabilire valutazioni su qualità, autenticità o affidabilità delle informazioni stesse (cfr. fig. 2). Ecco che un'attività, avvenuta in un certo momento, è associata ad un agente, ovvero anche una persona o un ente o un'istituzione, che è responsabile, nel corso di quella attività, della creazione di un'entità, ovvero di un oggetto qualificato.

---

<sup>8</sup> The PROV Data Model, <<https://www.w3.org/TR/2013/REC-prov-dm-20130430/>>.

<sup>9</sup> The PROV Ontology, <<http://www.w3.org/TR/prov-o/>>.

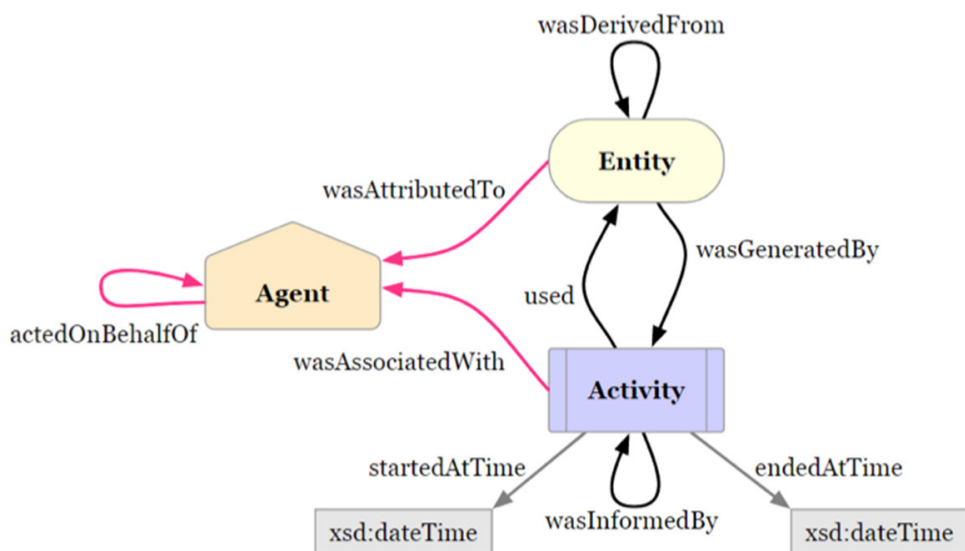


Fig. 2. Modello base di PROV-O.

The PROV Data Model

<<https://www.w3.org/TR/2013/REC-prov-dm-20130430/>>

Questo è il motivo per cui formalizzare attraverso modello ontologico il principio di provenienza, nei termini di un'azione svolta da un agente che opera sull'oggetto culturale come entità, è una garanzia di preservazione della conoscenza. Allo scopo di dare consistenza a questi ragionamenti non sarà privo di interesse provare ad esplorare il percorso che può condurre alla tutela del patrimonio culturale attraverso procedimenti ontologici. Partendo proprio dal modello suggerito da PROV-O e ricollocandolo nell'orizzonte culturale che ci interessa: quello della preservazione di un'azione intesa come un atto interpretativo condotto da un agente su un qualche aspetto del contenuto dell'oggetto culturale. Vale a dire garantire autorevolezza attraverso il contesto.

PROles<sup>10</sup> in particolare è un'ontologia nata originariamente per arricchire la rappresentazione di record di autorità in ambito archivistico (Daquino *et al.* 2014). L'obiettivo del modello, all'atto della sua ideazione, era la formalizzazione di relazioni complesse fra agenti ed eventi in cui questi agenti sono coinvolti quando ricoprono un certo ruolo politico, estraendo queste informazioni direttamente dal full-text di documenti testuali (ovvero anche gli oggetti culturali), che specificano il contesto dell'azione (i.e. persona-ruolo-evento come entità complessa attestata in una fonte). Questo approccio si inserisce in un discorso più ampio di approfondimento delle relazioni che possono intercorrere fra agenti (nella forma di enti, famiglie e persone secondo il modello EAC-CPF)<sup>11</sup> e documenti in cui questi agenti sono menzionati.

PROles, in particolare, importa ed estende due modelli: l'ontologia PRO<sup>12</sup> e l'Ontology Pattern Nary Participation.<sup>13</sup> Al di là degli aspetti tecnologici, PRO è stata concepita nell'ambito bibliografico per descrivere i ruoli degli agenti coinvolti nel processo di pubblicazione di un documento; in particolare, questa ontologia definisce una classe per specificare congiuntamente l'intervallo temporale e il contesto attorno ai quali viene esercitato un dato ruolo. Unitamente ad un primo livello di formalizzazione delle relazioni (descrivere un agente con un certo ruolo in un certo contesto), la descrizione delle informazioni che possono essere estratte dal full-text spesso richiede un ulteriore livello di contestualizzazione, ovvero la descrizione degli agenti che partecipano ad eventi, collocati nello spazio e nel tempo, ricoprendo un certo ruolo.

---

<sup>10</sup> Political Roles Ontology (PROles), <<http://www.essepuntato.it/2013/10/politicalroles>>.

<sup>11</sup> Encoded Archival Context - Corporate Bodies, Persons, and Families (EAC-CPF), <<http://eac.staatsbibliothek-berlin.de/>>.

<sup>12</sup> Publishing Roles Ontology (PRO), <<http://purl.org/spar/pro>>.

<sup>13</sup> Nary Participation, <[http://ontologydesignpatterns.org/wiki/Submissions:Nary\\_Participation](http://ontologydesignpatterns.org/wiki/Submissions:Nary_Participation)>.

lo. Per consentire questo ulteriore livello di descrizione, PRoles include anche Nary Participation ontological pattern,<sup>14</sup> che permette di modellare ogni oggetto nei termini della sua partecipazione ad un evento.

Un ulteriore livello di rappresentazione, quello che ci interessa più da vicino, è esattamente quello della definizione di responsabilità o *provenance* sulle affermazioni fatte in merito agli oggetti; la *provenance* è, in questo modello, un meta-livello necessario per attribuire la paternità delle asserzioni e legare le dichiarazioni fatte su agenti, ruoli ed eventi all'interprete che ha estratto queste informazioni dai documenti (cfr. fig. 3). La *provenance* è fondamentale sia per attribuire la paternità delle asserzioni fatte sui documenti (chi ha detto cosa), sia per individuare le responsabilità sulla creazione dei dati (chi ha formalizzato in tripla RDF qualcosa asserito da parte di qualcuno). Per questo aspetto di descrizione, PRoles importa alcuni predicati di PROV-O.

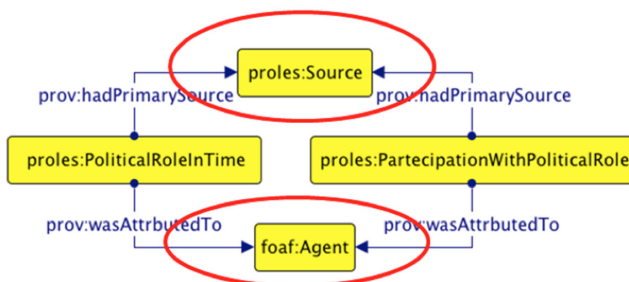


Fig. 3. Il terzo livello di PRoles: la dichiarazione di *provenance*.

Un agente dichiara che – in una fonte (un documento) – è attestato che un altro agente, che ha un ruolo specifico, ha partecipato a un determinato evento (modello di Silvio Peroni e Marilena Daquino)

<sup>14</sup> Lo scopo di un “ontological pattern” è di fornire un set di elementi ontologici che vogliono risolvere un problema specifico e che possono essere dunque riutilizzati in altre ontologie che si pongono la stessa tipologia di quesiti (cfr. per esempio Gangemi 2005). In particolare Nary Participation vuole fornire gli strumenti ontologici necessari a rispondere alle domande: chi sono i partecipanti ad un dato evento in un certo momento? Quale evento ha visto coinvolti quali partecipanti in un dato luogo?

Partendo dal problema della *provenance*, come formalizzato in PRoles, si è riflettuto sull'esigenza di estendere l'analisi alle entità coinvolte nel processo interpretativo, per arrivare a stabilire, secondo criteri condivisibili, quali asserzioni possano essere considerate autorevoli e/o meglio documentate rispetto ad un'asserzione priva di contesto che ne avvalori il contenuto.

HiCO, "Historical Context Ontology"<sup>15</sup> (Daquino - Tomasi 2015), è una ontologia nata esattamente per descrivere il contesto di oggetti culturali in quanto prodotto del *workflow* interpretativo di un agente. Questo modello consente di gestire affermazioni autorevoli su asserzioni, potenzialmente anche contraddittorie, prevedendo meccanismi che possano consentire la coesistenza di interpretazioni esito di processi analitici difformi. Ogni asserto è concepito infatti come una lettura soggettiva di un interprete su uno specifico livello della fonte da cui tale lettura viene estratta, come ad esempio un testo, o meglio l'«espressione» del testo della fonte (cfr. fig. 4). Un testo è per sua natura un oggetto multi-livellare: HiCO, acquisendo il modello FRBR,<sup>16</sup> cerca di far fronte a questa complessità per determinare i livelli della rappresentazione dell'oggetto informativo (l'opera, la sua espressione, le diverse possibili manifestazioni e l'item).

In particolare HiCO – che riusa FRBR, PRO, Nary e PROV-O – permette di descrivere l'atto interpretativo di un agente (`hico:InterpretationAct` class) sulla base di 'tipologie' di interpretazioni (`hico:hasInterpretationType` property) e 'criteri' interpretativi (`hico:hasInterpretationCriterion` property), e, attraverso l'ontologia CiTO,<sup>17</sup> consente di creare relazioni fra interpretazioni riferite allo stesso oggetto culturale fatte da autori/editori/commentatori diversi.

---

<sup>15</sup> HiCO (Historical Context Ontology), <<http://purl.org/emmedi/hico>>

<sup>16</sup> FRBR, [purl.org/spar/frbr](http://purl.org/spar/frbr); *Guidelines*, <<http://www.ifla.org/publications/functional-requirements-for-bibliographic-records>>

<sup>17</sup> CiTO (Citation Typing Ontology), <<http://purl.org/spar/cito/>>

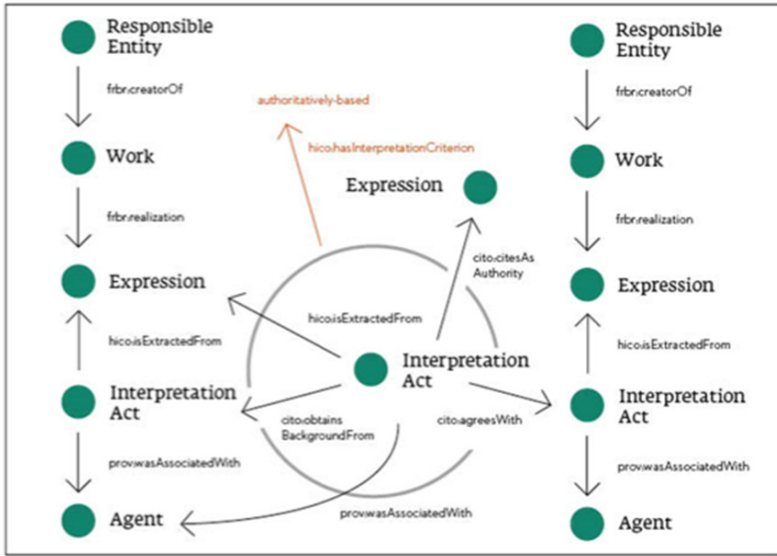


Fig. 4 Il modello HiCO.

A sinistra l'oggetto nativo da descrivere, a destra un altro oggetto culturale, espressione dell'interpretazione dell'oggetto nativo da parte di un agente (modello di Marilena Daquino).

### *LOD e preservazione della conoscenza*

L'impiego di modelli ontologici per la formalizzazione del sapere non può funzionare se tali modelli non vengono testati, ovvero se la loro capacità espressiva non viene verificata attraverso l'applicazione della concettualizzazione ad una realtà di oggetti culturali esistenti.

L'impiego di HiCO come strumento per formalizzare le interpretazioni è stato quindi testato su due *datasets*: un archivio fotografico e un'edizione di lettere. L'ontologia è stata quindi impiegata per affrontare due opposti metodi di rappresentazione del contenuto di un oggetto culturale: quello data-centrico tipico dei sistemi di descrizione catalografica e quello documento-centrico tipico delle *scholarly editions*. Se quindi nel primo caso lo sforzo di traduzione in LOD si deve concentrare sul processo di identificazione univoca degli elementi atomici e di conversione dei campi descrittivi in classi e proprietà, nel secondo caso



il documento deve essere analizzato per estrarre stringhe sotto forma di dati strutturati. Ma in entrambi i casi è necessario legare il processo interpretativo, ovvero la compilazione dei campi della scheda da un lato e il markup dall'altro, all'agente che ha formulato l'interpretazione, e legare l'interpretazione alla fonte che la supporta. Ovviamente l'agente che ha provveduto alle informazioni descrittive dell'oggetto non necessariamente sarà il creatore della tripla. Questo significa che la *provenance* dovrà garantire l'adeguata documentazione di ogni agente coinvolto nel processo ermeneutico: l'interprete della fonte, l'autore della fonte, il creatore/interprete della tripla.

Il primo caso, si riferisce all'esperienza condotta sull'archivio fotografico Zeri,<sup>18</sup> un catalogo ad oggi considerato tra i più imponenti repertori di arte italiana sul Web. Il progetto completo<sup>19</sup> ha previsto, assieme alla modellazione ontologica degli standard di descrizione catalografica Scheda F e Scheda OA,<sup>20</sup> la produzione e pubblicazione del dataset in RDF, la creazione di collegamenti al LOD cloud e alle più opportune authorities, e la definizione di un'interfaccia per il browsing (cfr. in particolare Gonano *et al.* 2014; Daquino *et al.* 2017a; Daquino *et al.* 2017b).

La partecipazione della Fondazione al progetto Pharos (Reist *et al.* 2015), un consorzio internazionale di archivi fotografici, permetterà di sperimentare l'uso di HiCO per confrontare le diverse attività, condotte dai catalogatori delle istituzioni coinvolte, nella descrizione delle medesime risorse (per esempio l'attribuzione di paternità ad un'opera d'arte rappresentata in una fotografia) e trarne quindi delle inferenze. Attraverso regole SWRL applicate ai rapporti tra le fonti, ai criteri e alle informazioni di contesto utilizzati da un agente per spiegare la sua in-

---

<sup>18</sup> Catalogo della Fondazione Zeri, <<http://www.fondazionezeri.unibo.it/it>>.

<sup>19</sup> Zeri & LOD. Dataset e documentazione, <<http://data.fondazionezeri.unibo.it/>>.

<sup>20</sup> FEntry Ontology, <<http://www.essepuntato.it/2014/03/fentry>>; OAEntry Ontology, <<http://purl.org/emmedi/oaentry>>.

interpretazione, potremmo infatti dedurre formalmente quando una interpretazione può essere considerata autorevole (Daquino *et al.* 2016).

Il secondo caso è l'edizione digitale delle lettere, inviate/ricevute a/dal copista fiorentino Vespasiano da Bisticci, vissuto nell'arco del XV secolo<sup>21</sup> (Tomasi 2013).

In questo caso al testo delle lettere, come trascrizione diretta delle fonti primarie in formato XML/TEI, sono aggiunti una serie di commenti non strutturati che sono stati tradotti in RDF, istituendo una relazione fra il testo della lettera ed il commento effettuato dall'editore. Attraverso l'ontologia CITO è stato possibile stabilire il rapporto fra testo e commento e fra commento e fonti. L'editore, nell'arricchire il testo con informazioni storiche, prosopografiche, codicologiche e culturali in senso estensivo, assume posizioni sulla base tanto delle proprie conoscenze quanto alla luce della consultazione di fonti. L'asserto lega dunque l'attribuzione di paternità dell'interpretazione alla fonte che corrobora tale posizione o anche alla fonte che la confuta o ancora che dichiara posizioni diverse. La tripla RDF, prodotta da un altro agente, esprime l'affermazione dell'editore e il suo rapporto con la fonte, in relazione al contenuto dell'oggetto descritto.

### *Conclusioni*

Preservare è dunque un'azione composita, che affronta la complessità del concetto di oggetto digitale attraverso il principio della multi-livellarità. La *provenance* come dichiarazione della storia dell'oggetto culturale è fondamentale per stabilire un primo livello di affidabilità degli asserti legati agli oggetti del patrimonio. Ed è dunque necessaria una sua concettualizzazione, ovvero la sua conversione in un modello ontologico, per garantire la preservazione di informazioni necessarie alla tutela della memoria.

---

<sup>21</sup> Vespasiano da Bisticci, *Lettere*, <<http://vespasianodabisticciletters.unibo.it/>>.

Nel nostro ultimo esempio, i dati della ricerca, rappresentati dalle schede di descrizione catalografica e dal markup del full-text, vengono preservati logicamente attraverso lo schema (schema del database e schema di markup) e concettualmente attraverso l'ontologia. Attraverso la costruzione di opportuni collegamenti fra l'interpretazione e il suo responsabile, ma anche fra l'interpretazione e la fonte utilizzata per dichiarare quella interpretazione, è possibile mettere in relazione interpretazioni diverse, per stabilire una tassonomia di casi, contemplando la contraddittorietà di asserti adeguatamente documentati. Questo significa anche mettere in condizione gli utenti di scegliere quale interpretazione ritenere più autorevole, in base all'attribuzione di paternità delle operazioni. La fiducia di un utente A è veicolata certo dalla conoscenza diretta di utente B, ma anche dalla capacità dello stesso utente B di documentare adeguatamente il contesto di un asserto, per esempio attraverso esplicita menzione delle fonti. Un avvicinamento importante al *Web of Trust* quindi, per contrassegnare le asserzioni RDF al fine di stabilirne la provenienza e quindi garantirne autenticità e affidabilità, ma anche muoversi nella prospettiva dell'autorevolezza.

Mentre l'uso di metadati consente la preservazione del dato e, attraverso lo schema logico, del contenuto informativo, con le ontologie la preservazione mira a garantire l'accesso alla conoscenza, attraverso il modello concettuale che attribuisce al metadato lo spessore semantico necessario a correlare il dato al suo contesto.

La formalizzazione della *provenance*, come principio che l'archivistica ha assunto come elemento determinante del concetto di contesto, avvalorata il principio di autenticità, fornendo all'utente finale un dato potenzialmente affidabile. Ma al contempo il modello ontologico permette la coesistenza di interpretazioni difformi, fornendo strumenti per valutare però l'autorevolezza degli asserti ed effettuare di conseguenza scelte consapevoli.

## Bibliografia

- Abele *et al.* 2017 = Andrejs Abele - John P. McCrae - Paul Buitelaar - Anja Jentzsch - Richard Cyganiak, *Linking Open Data cloud diagram*, 2017 <<http://lod-cloud.net/>>.
- Assman 1999 [2002] = Aleida Assmann, *Erinnerungsräume. Formen und Wandlungen des kulturellen Gedächtnisses*, München, C. H. Beck, 1999; trad. it. *Ricordare. Forme e mutamenti della memoria culturale*, Bologna, il Mulino, 2002.
- Corrado - Sandy 2004 = Edward M. Corrado - Heather Moulaison Sandy, *Digital Preservation for Libraries, Archives, and Museums*, Maryland (USA), Rowman & Littlefield, 2004.
- Daquino *et al.* 2014 = Marilena Daquino - Silvio Peroni - Francesca Tomasi - Fabio Vitali, *Political Roles Ontology (PROles): enhancing archival authority records through Semantic Web technologies*, «Procedia Computer Science», 38 (2014), p. 60-67.
- Daquino - Tomasi 2015 = Marilena Daquino - Francesca Tomasi, *Historical Context (HiCO): a conceptual model for describing context information of cultural heritage objects*, in *Metadata and Semantics Research*, edited by Emmanouel Garoufallou, Richard J. Hartley, Panorea Gaitanou, Cham etc., Berlin, Springer, 2015, p. 424-436.
- Daquino *et al.* 2017a = Marilena Daquino - Silvio Peroni - Francesca Tomasi - Fabio Vitali, *The Project Zeri Photo Archive: Towards a Model for Defining Authoritative Authorship Attributions*, in *Digital Humanities 2016: Conference Abstracts*, Kraków, Jagiellonian University & Pedagogical University, 2016, p. 472-474.
- Daquino *et al.* 2017b = Marilena Daquino - Francesca Mambelli - Silvio Peroni - Francesca Tomasi - Fabio Vitali, *Enhancing Semantic Expressivity in the Cultural Heritage Domain: Exposing the Zeri Photo Archive as Linked Open Data*, «ACM Journal of

- Computing and Cultural Heritage»,10 (July 2017), 4, article 21 <<http://dx.doi.org/10.1145/3051487>>.
- DCH-RP 2013 = *A Roadmap for preservation of digital cultural heritage content*, edited by Börje Justrell (Riksarkivet) and Antonella Fresa (Promoter Srl), 2013 <[http://www.digitalmeetsculture.net/wp-content/uploads/2013/10/Handbook\\_A\\_Roadmap\\_for\\_preservation\\_for\\_digital\\_cultural\\_heritage\\_content.pdf](http://www.digitalmeetsculture.net/wp-content/uploads/2013/10/Handbook_A_Roadmap_for_preservation_for_digital_cultural_heritage_content.pdf)>.
- Deegan - Tanner 2006 = *Digital Preservation*, edited by Marilyn Deegan - Simon Tanner, London, Facet, 2006.
- Digital Preservation Coalition 2015 = Digital Preservation Coalition, *Digital Preservation Handbook*, 2nd edition, 2015 <<http://handbook.dpconline.org/>>.
- Gangemi 2005 = Aldo Gangemi, *Ontology Design Patterns for Semantic Web Content*, in *Proceedings of the Fourth International Semantic Web Conference*, edited by Mark A. Musen *et al.*, Heidelberg, Springer, 2005
- Lavoie - Gartner 2013 = Brian Lavoie - Richard Gartner, *Preservation Metadata: Digital Preservation Coalition Technology Watch Report 13-03*, 2nd edition, 2013 <<http://dx.doi.org/10.7207/twr13-03>>.
- Garret - Waters 1996 = John Garrett - Donald Waters, *Preserving Digital Information. Report of the Task Force on Archiving of Digital Information*, Washington, DC, CLIR, May 1996 <<http://www.clir.org/pubs/reports/pub63watersgarrett.pdf>>.
- Gonano *et al.* 2014 = Ciro Mattia Gonano - Francesca Mambelli - Silvio Peroni - Francesca Tomasi - Fabio Vitali, *Zeri e LOD. Extracting the Zeri photo archive to Linked Open Data: formalizing the conceptual model*, in *Digital Libraries (JCDL)*, London, IEEE, 2014, p. 289-298.
- Higgins 2011 = Sarah Higgins, *Digital Curation: The Emergence of a New Discipline*, «The International Journal of Digital Curation», 6 (2011), 2, p. 78-88 <<http://ijdc.net/index.php/ijdc/article/view/184>>.

- OAIS 2012 = ISO 14721 *Space data and information transfer systems - Open archival information system - Reference model*, 2012 <<https://public.ccsds.org/publications/archive/652x1m2.pdf>>.
- Pratesi 1979 = Alessandro Pratesi, *Genesi e forme del documento medievale*, Roma, Jouvence, 1979.
- Premis 2008 = PREMIS Editorial Committee, *PREMIS Data Dictionary for Preservation Metadata*, version 2.0, March 2008 <<http://www.loc.gov/standards/premis/v2/premis-2-0.pdf>>.
- Reist *et al.* 2015 = Inge Reist - David Farneth - Samuel Stein - Remigius Weda, *An Introduction to PHAROS: Aggregating Free Access to 31 Million Digitized Images and Counting*, Paper presented at CIDOC 2015, New Dehli, India, September 5-10, 2015.
- Ross 2012 = Seamus Ross, *Digital Preservation, Archival Science and Methodological Foundations for Digital Libraries*, «New Review of Information Networking», 17 (2012), 1, p. 43-68.
- Testoni 2013 = Laura Testoni, *Digital curation e content curation: due risposte alla complessità dell'infosfera digitale che ci circonda, due sfide per i bibliotecari*, «Bibliotime», XVI (marzo 2013), 1.
- Thibodeau 2002 = Kenneth Thibodeau, *The State of Digital Preservation: An International Perspective*, in *Overview of Technological Approaches to Digital Preservation and Challenges in Coming Years*, CLIR, 2002, p. 4-31.
- Tomasi 2010 = Francesca Tomasi, *Rappresentare e conservare*, in Domenico Fiormonte - Teresa Numerico - Francesca Tomasi, *L'umanista digitale*, Bologna, Il Mulino, 2010, p. 119-164.
- Tomasi 2013 = Francesca Tomasi, *L'edizione digitale e la rappresentazione della conoscenza. Un esempio: Vespasiano da Bisticci e le sue lettere*, «Ecdotica», 9 (2013), p. 264-286.
- Webb 2003 = Colin Webb, *Guidelines for the preservation of digital heritage*, Information Society Division United Nations Educational, Scientific and Cultural Organization, 2003 <<http://unesdoc.unesco.org/images/0013/001300/130071e.pdf>>.

## Abstract

La preservazione è un'attività composta, in virtù del fatto che gli oggetti digitali sono entità multilivello, stratificate e complesse. Per garantire la preservazione concettuale è necessario che l'oggetto culturale sia corredato delle più puntuali informazioni relative alla provenance. Con la provenance è infatti possibile attribuire paternità alle azioni legate a quell'oggetto culturale, in modo tale da garantirne l'affidabilità. Vista la quantità di informazioni esistenti sul Web vanno previsti meccanismi utili a stabilire l'origine dei dati, per garantirne oltre all'affidabilità anche l'autorevolezza, affinché il fruitore finale se ne possa avvalere. Con la traduzione di questi principi in ontologia si garantisce che l'oggetto venga preservato nella sua dimensione concettuale. Vengono allora descritte due ontologie finalizzate alla formalizzazione della provenance (PROles e HiCO), che sono state anche testate su due collezioni di oggetti culturali: una collezione strutturata (un catalogo) e una raccolta di documenti non strutturati (un'edizione digitale).

Ontologie, preservazione, provenienza, atto interpretativo

*Preservation is a composite task, because digital objects are multilevel, stratified, and complex entities. In order to ensure conceptual preservation, the cultural object must be accompanied by the most accurate details about the provenance. With the provenance, in fact, it is possible to attribute paternity to the actions linked to that cultural object, in such a way as to guarantee its reliability. Given the amount of existing information on the Web, there is a need for mechanisms to determine the provenance of the data, in order to guarantee not just the reliability of the information but also the authoritativeness, so that the final user can use it. By translating these principles into ontology,*

*the object is preserved in its conceptual dimension. Two ontologies aimed at the formalization of the provenance (PROles and HiCO) are then described, which have also been tested on two different projects: a structured collection (a catalog) and a set of unstructured documents (a digital edition).*

*Ontologies, preservation, provenance, interpretation act*