



Manifestazioni-dopo

Riflessioni sulla *Digital Preservation*

Daniela Canali <daniela.canali@tin.it>

ISPRI-CNR, Istituto di Studi Socio-Economici sull'Innovazione e le Politiche della Ricerca - Consiglio Nazionale delle Ricerche, Roma

Le presenti riflessioni scaturiscono dalla partecipazione al seminario *Preserving the Record of Science*, svoltosi nel febbraio scorso a Parigi presso la sede dell'UNESCO, organizzato da ICSTI [1], CODATA [2] e ICSU [3], che ha radunato oltre 80 partecipanti da differenti organizzazioni e Paesi, in rappresentanza di interessi diversi nel campo della scienza. Nel gennaio del 2000 l'ICSTI aveva già organizzato un seminario sullo stesso tema, spinto dalla crescente produzione elettronica di pubblicazioni scientifiche, paventando il rischio di mancata archiviazione delle informazioni.

Tema ricorrente del *meeting* è stato il necessario incoraggiamento diretto agli scienziati a comprendere l'importanza dell'archiviazione dei propri contributi e a rendersi conto dell'effettivo valore della raccolta che tale procedura consente di creare. C'è stata una notevole discussione anche sui vari ruoli richiesti dall'attività di *preservation* e su quali persone e istituzioni debbano farsi carico di tale responsabilità. Opinioni differenti sono poi emerse circa la selezione dei dati da conservare: la distinzione tra memoria e conservazione, applicate ai risultati di attività scientifiche, necessita di ulteriore chiarimento, mentre è apparso chiaro da tempo che l'archiviazione digitale è un'attività costosa e che gli enti finanziatori hanno diritto di chiedere giustificazione dei costi.

1. Metadata

Il tema dei metadati, i dati estratti dai documenti per consentirne l'identificazione e l'accesso [4], è ritornato più volte al centro dei dibattiti, in relazione a varie questioni: in molti dei progetti di *digital archiving* in corso il Dublin Core, il *set* di metadati concordato dalla comunità bibliografica tradizionale, è usato come standard *ad hoc*, mentre molte organizzazioni lo accettano solo come fondamento,

considerando validi altri schemi possibili.

Partecipava ai lavori un *working group* sui *preservation metadata* [5], costituito dai rappresentanti di organizzazioni con responsabilità di *archiving* che da tempo discutono i problemi legati all'uso dei metadata per le loro necessità. In particolare, la comunità rappresentata da CODATA ha espresso la sensazione che molte iniziative sui *metadata* siano più orientate al testo che ai dati, come vengono intesi nella loro comunità. Di conseguenza, si è giunti all'accordo che un gruppo misto di rappresentanti ICSTI e CODATA, col supporto del proposto CODATA Task Group su *preservation* e *archiving* di dati scientifici e tecnici nei Paesi in via di sviluppo, controllerà le attività del Working Group sui *preservation metadata* con uno sguardo all'identificazione di problemi comuni ed in particolare alle esigenze emergenti da dibattere nell'ambito del Working Group o, se necessario, più ampiamente. Infatti la prossima conferenza CODATA che si terrà a Montreal affronterà il tema della *digital preservation* e sarà un'occasione per analizzare altri aspetti di questo argomento.

Il cambiamento radicale verso documenti digitali originali e la stessa disseminazione digitale dei risultati delle attività scientifiche fa sì che la distinzione storica tra testo e dati stia scomparendo. Materiali "born digital" sono *bit* e *byte* e possono essere trattati esattamente allo stesso modo, a prescindere dal significato che veicolano. La discussione sui metadata in generale deve ampliarsi fino a tener conto delle necessità di *archiving*: ciò si può applicare in particolare alle discussioni relative all'Open Archive Initiative. In tale contesto diventa di primaria importanza definire le modalità di creazione e di mantenimento di identificatori interoperabili permanenti e la loro relazione con l'accesso agli archivi. Il lavoro del gruppo Digital Object Identifier (DOI) è da tener presente in ogni pianificazione futura.

2. Modelli economici

Dalle discussioni sono emersi diversi problemi relativi alla struttura economica delle attività di *digital preservation*, dal momento che per le iniziative attualmente in corso non c'è un valido modello economico o di *business*, questione che ha invece un peso tale da determinare la continuità o meno dell'archivio. I progetti, specialmente quelli che mirano a creare archivi di significativa entità, dovrebbero essere avviati solo a fronte di una rigorosa giustificazione economica del loro valore, anche per assicurarne la continuità.

Il progetto europeo di *digital preservation* recentemente finanziato, ERPANET - European Consortium for Digital Preservation, apporta un nuovo *input*: il progetto intende monitorare casi studio di azioni di *preservation* focalizzati sull'estrazione della conoscenza del valore a favore di altri progetti [6], attività che fornisce notevole introspezione anche nell'area economica.

In un contesto più ampio l'esperienza nello sviluppo di nuove azioni in ambito digitale, come lo sviluppo di PubliMed, fornisce puntatori e coniuga il punto di vista di studiosi, editori e biblioteche, di tutti coloro, cioè, che hanno interesse negli aspetti economici dell'iniziativa, come pure nel valore generale e nelle esigenze dell'*archiving*.

STM, l'associazione che rappresenta editori scientifici, tecnici e medici, e ALPSP, l'associazione di Learned & Professional Society Publishers, sono impegnate ad esaminare in dettaglio gli aspetti economici della *digital preservation*: dovrebbero cominciare tentando di raccogliere dai loro associati dati sulle future azioni e i diversi punti di vista circa gli elementi necessari a sviluppare validi modelli di *business*, come pure i risultati delle attività già intraprese. Nel selezionare i casi-studio il progetto ERPANET tiene ben presenti i problemi economici ed è aperto ai suggerimenti.

3. Migrazione tecnologica

Chi si occupa di archiviazione digitale deve fare i conti con la necessità di assicurare l'accesso futuro all'archivio al di là dei cambiamenti tecnologici che si vengono a verificare.

Esiste già materiale a tutti gli effetti perduto in quanto inaccessibile, dal momento che non sono più disponibili dispositivi per leggerlo. La situazione è probabilmente destinata a peggiorare con la velocità dei cambiamenti tecnologici: per permettere la migrazione tecnologica diverse iniziative devono essere intraprese, ma ci sono questioni aperte sulla loro validità e sulla gamma di opzioni da prendere in considerazione (per esempio, l'emulazione).

Le organizzazioni con interessi nel campo più ampio dell'ICT (Information and Communication Technologies) devono essere coinvolte nell'affrontare tale questione, per identificare aspetti di comune interesse: a questo proposito, è stata citata una nuova legislazione che richieda ai costruttori di apparecchiature nel settore dell'ITC di tener traccia delle loro posizioni a scopi di riciclaggio; la gestione del bene digitale è di fatto un'attività in via di sviluppo e le tecnologie di *storage* per gli oggetti digitali un campo in rapida espansione.

Sebbene non si sia giunti ad alcun accordo, è chiaro che le organizzazioni con specifiche funzioni nel campo dell'*archiving*, quali biblioteche nazionali e centri produttori di dati scientifici, hanno la responsabilità di mantenere memoria in tale area, essendo direttamente interessate. Anche la possibilità di sviluppare "test beds" per le valutazioni è considerata una priorità importante.

4. Relazioni con la comunità archivistica

Gli archivisti professionisti si sono già confrontati con molti dei problemi ben noti a coloro che sono coinvolti nella *digital preservation*; in particolare, azioni relative alla selezione e alla valutazione di materiali destinati all'archiviazione, problemi di privacy e di confidenzialità (di primaria importanza nelle scienze sociali) e la migrazione tecnologica.

I rappresentanti della comunità archivistica presenti al *meeting* hanno deciso di trasmettere all'ICSTI, per una più ampia circolazione, i risultati di varie discussioni, liste e gruppi di lavoro rilevanti per gli interessi della *digital preservation*. L'ICA, International Council of Archivists, ha già gruppi di lavoro e altri organismi impegnati nella *preservation* di *record* di scienza e tecnologia, che operano a stretto contatto con la comunità di storici della scienza.

Organizzazioni scientifiche autonome sono state incoraggiate ad identificare rappresentanti che possano partecipare e collaborare agli sviluppi in questo settore.

5. Standard

Anche il tema degli standard è emerso ripetutamente. Azioni formali, quali la posizione dell'ISO sull'Open Archival Information System (OAIS) Reference Model, vanno affrontate a breve, come pure l'attività nella comunità OAI relativa al *metadata harvesting*. Attualmente c'è viva discussione intorno al tema degli standard *ad-hoc* e agli accordi sulla possibilità di seguire schemi particolari.

Sebbene sia chiaro che diverse comunità stanno lavorando alla definizione di standard in funzione di problemi specifici, viene considerato improbabile e persino inutile forzare un coordinamento e un'adesione su larga scala ad un insieme comune di norme. Comunque, i flussi informativi sulle attività dei vari gruppi di lavoro e di altri organismi che si occupano di standard, o attività affini, sono di fondamentale importanza e maggiori sforzi per migliorare la circolazione dell'informazione sono sicuramente apprezzabili.

Lo scambio tra coloro che sono storicamente interessati ai dati scientifici, gli archivisti e la comunità del testo va senz'altro incoraggiato. I partecipanti al *meeting* si sono impegnati ad assicurare che la loro comunità venga messa al corrente di tali problemi, come punto di partenza, in alcuni casi, per il miglioramento dei flussi di informazione. Le attuali azioni di disseminazione della conoscenza, come la Digital Preservation Coalition, supportata da JISC in Gran Bretagna, e la notevole attività di supporto a livello di Unione Europea, andrebbero estese ad altre comunità e ad altri Paesi.

6. Paesi in via di sviluppo

Il problema della *digital preservation* non è ovviamente limitato alle nazioni sviluppate. I Paesi in via di sviluppo, in particolare Asia del Sud e Sud-Est e America Latina, sono produttori significativi di informazione scientifica. La loro esigenza di preservare in formato digitale il materiale prodotto è la medesima dei Paesi sviluppati.

Il Task Group di CODATA terrà un workshop in Sud Africa nel maggio 2002, al quale parteciperanno i Paesi del Southern African Development Council. Il gruppo CODATA intende ripetere tali iniziative anche in altre regioni in via di sviluppo. Al *meeting* del Governing Body nel novembre 2001 l'UNESCO ha aggiunto alle azioni previste sulla preservazione del patrimonio culturale un'azione per includere la *digital preservation*.

Anna Maria Cetto, rappresentante di Latindex, iniziativa che raggruppa editori scientifici e biblioteche in America Latina, presente al seminario, si è impegnata ad identificare organizzazioni e persone, in Paesi e regioni in via di sviluppo, che abbiano interesse nella *digital preservation*, per metterli al corrente dei risultati del seminario ed incoraggiarli a partecipare ad azioni rilevanti per le loro necessità.

7. Training

Il *training* su aspetti specifici delle procedure per la *digital preservation*, come pure l'educazione al valore della creazione di archivi, sono stati argomenti di discussione nel *meeting*. Un'azione specifica di *training* è stata inclusa nei programmi UNESCO per i Paesi e le regioni in via di sviluppo. L'esperienza raggiunta da parte di alcuni editori nelle iniziative in corso di *digital preservation* potrebbe essere condivisa con coloro che non hanno ancora affrontato il problema. Inoltre, i casi studio del progetto ERPANET dovrebbero costituire prove valutabili in questo ambito.

Considerazioni conclusive

A livello generale la preservazione di informazione scientifica digitale originale è riconosciuta come un'attività significativamente in espansione: rispetto alla situazione al tempo del precedente meeting ICSTI nel 2000, sono state intraprese parecchie iniziative e ben maggiore è il coinvolgimento di un alto numero di organizzazioni.

I diversi settori delle scienze stanno ora utilizzando un mezzo comune, il contenuto digitale, per la produzione e la disseminazione dei loro risultati, il che significa che i problemi affrontati nella preservazione dei risultati sono comuni: in questo senso emergono i vantaggi nella condivisione dell'esperienza e nello sviluppo di approcci comuni. Una notevole mole di lavoro è ancora da fare nel portare alcune realtà al livello di progresso attuale e va attentamente considerata l'opportunità di ampliare le conoscenze sulla digital preservation imparando da aree esterne a quella formale della ricerca scientifica e tecnologica, come la comunità archivistica, le comunità archeologiche e di scienze sociali e le società di ICT.

Il seminario ha rappresentato un valido punto di incontro per una pluralità di interessi e una buona occasione per creare comunicazione e per valutare l'opportunità di azioni comuni. Altri *meeting* verranno organizzati nei mesi venturi per ampliare il flusso di informazione e l'ICSTI, con CODATA e ICSU, continuerà a monitorare gli sviluppi nell'interesse della scienza e degli studiosi.

Note

[1] L'International Council for Scientific and Technical Information costituisce un forum unico per l'interazione tra le organizzazioni che creano, disseminano ed usano informazioni tecnico-scientifiche. La sua missione è relativa alle discipline scientifiche e tecniche a livello internazionale, per offrire alle organizzazioni che ne fanno parte una comunità globale.

[2] Il Committee on Data for Science and Technology, è un comitato scientifico interdisciplinare dell'International Council for Science (ICSU): ha l'obiettivo di migliorare la qualità e l'accessibilità dei dati, oltre che dei metodi con cui essi vengono acquisiti, gestiti, analizzati e valutati con particolare attenzione ai Paesi in via di sviluppo. Cerca inoltre di agevolare la cooperazione internazionale tra coloro

che raccolgono, organizzano e usano i dati e di promuovere una crescente consapevolezza nella comunità tecnica e scientifica dell'importanza di tali attività, accanto alla considerazione dei problemi legati all'accesso e alla proprietà intellettuale.

[3] L'International Council for Science mira a rompere le barriere della specializzazione avviando e coordinando un maggior numero di programmi di interdisciplinarietà internazionale e creando istituzioni interdisciplinari che assicurino attività e programmi di ricerca di interesse per i diversi membri. Un certo numero di organizzazioni operanti all'interno dell'ICSU affronta inoltre problemi comuni a tutti gli studiosi, quali la capacità di costruire nella scienza, l'ambiente e lo sviluppo e la libertà scientifica. Il Council opera come *focus* per lo scambio di idee e informazioni e per lo sviluppo di standard.

[4] Il termine "metadata" è utilizzato in una grande varietà di contesti, dall'uso nelle funzioni di controllo delle operazioni dei DBMS a quello per la descrizione di *dataset* scientifici e di supporto alla condivisione dei dati tra gli studiosi, fino ai *metadata* usati nelle *digital library* per supportare l'accesso dell'utenza all'informazione.

Impiegato nel contesto delle biblioteche, tradizionali o digitali, il concetto di *metadata* si riferisce di solito all'informazione che fornisce una caratterizzazione (di solito breve) del singolo *information object* nelle collezioni di una biblioteca; è memorizzata principalmente come contenuto dei cataloghi di biblioteca nelle biblioteche tradizionali; è usata principalmente per assistere gli utenti nell'accesso all'*information object* di interesse.

[5] Diverse istituzioni hanno sviluppato "element sets" di preservation metadata, dettagliando le informazioni specifiche che dovrebbero essere incluse nei *metadata* di un oggetto archiviato. In particolare, la pubblicazione dell'OCLC Working Group on Preservation metadata (31 gennaio 2001) analizza in modo comparativo quattro approcci d'uso, autonomamente sviluppati: CURL Exemplars in Digital Archives project (CEDARS), la National Library of Australia, la Networked European Deposit Library (NEDLIB) e l'Harvard University's Digital Repository Services (quest'ultimo è l'unico che non si ispira al modello OAIS).

[6] ERPANET <<http://www.20erpanet.org/>> è un consorzio europeo estensibile, nato con lo scopo di rendere disponibili informazioni ed esperienze e di migliorare le capacità nel campo della conservazione del patrimonio digitale culturale e scientifico. Il consorzio riunisce istituti della conservazione e della memoria (musei, biblioteche e archivi), industria di tecnologia informatica e telematica, centri di ricerca, organizzazioni governative (incluse quelle locali), industrie dell'intrattenimento ed il mondo della produzione.

ERPANET fornisce dunque una stanza di compensazione virtuale che offre informazioni e conoscenze sullo stato dell'arte e sugli sviluppi delle attività di conservazione digitale.