La conservazione nel Cloud

LUCIANA DURANTI
Chair and Professor, Archival Studies
The University of British Columbia, Vancouver, B.C. Canada
luciana.duranti@ubc.ca

Avremo un sistema di conservazione affidabile in futuro?

William Lehr scrive che oggi l'Internet è una "infrastruttura essenziale". Questa espressione fa riferimento al ruolo socio-economico della rete, simile a quello originariamente fornito dell'accesso universale alla telefonia. Considerando che l'*Internet Cloud* attiva un set di servizi in-rete esteso a tutti, come l'accesso a risorse digitali, archivi online, e altri servizi di più alto livello, oltre al trasporto di dati,[1] si può affermare, con Blanchette, che il Cloud è diventato un tipo di *meta-infrastruttura,* capace di una crescita sostenibile senza precedenti,[2] dove il termine infrastruttura è definito come un ecosistema computerizzato che fornisce servizi *alle applicazioni*, piuttosto che come applicazioni che forniscono *servizi agli utenti*.[3] Questo è il motivo per cui molti paesi cominciano a considerare il Cloud un'infrastruttura critica, cioè un'infrastruttura che ha una funzione vitale per l'economia e la società. È dunque logico aspettarsi che, in futuro, i sistemi di conservazione siano sempre più spesso nel Cloud piuttosto che all'interno di organizzazioni e istituti. Se essi saranno "affidabili", o se saranno sistemi, piuttosto che vari *agglomerati* di servizi basati su contratti con vari fornitori, dipenderà dall'abilità degli archivisti di sviluppare standard per una infrastruttura Cloud internazionale, e dal loro impatto sulle politiche governative e sull'opinione pubblica.

Luciana Duranti ha inviato questo contributo al seminario "La memoria fra le nuvole: di bit in bit, dal presente al futuro. La conservazione del digitale, i nuovi tipi di 'Beni culturali'", seminario MAB - Regione Lombardia, Milano, 17 Marzo 2016, Palazzo delle Stelline. Una versione estesa dell'articolo è pubblicata in inglese nel libro *Trustworthy Systems for Digital Objects: Theory and Practice*, a cura di Philip Bantin, Lanham, MD, Rowman & Littlefield Publishing Group, 2016.

Il Cloud

Non c'è un accordo sulla definizione di Cloud computing; solo un riconoscimento che si tratta di un modello che comprende una varietà di servizi collegati da una rete capillare, accessibili ovunque a più utenti, indipendentemente dalla collocazione dell'utente e dei fornitori di servizi (provider), offerti su richiesta e pagati proporzionalmente all'utilizzo. Tuttavia, questo modello può essere modificato in base alle necessità, per esempio offrendo un servizio solo ad un determinato utente, in un singolo posto, fuori dalla rete, su prenotazione, o con tariffa fissa. Infatti, Weinman crede che un approccio ibrido sia il miglior approccio all'utilizzo dei servizi del Cloud.[4] Anche l'Istituto degli standard e tecnologie degli Stati Uniti (NIST) considera il Cloud computing un "paradigma in evoluzione" che permette di combinare vecchie e nuove tecnologie in diverse maniere.[5]
Ci sono molte ragioni per cui gli archivisti sono restii a collocare l'immagazzinamento e la conservazione di dati in un ambiente Cloud. Queste ragioni sono collegate all'affidabilità e alla trasparenza dei servizi, alla sicurezza, alla privacy, al controllo e alla giurisdizione.[6] Molto è stato scritto su questo, tanto quanto sui vantaggi del Cloud, per la maggior parte collegati all'accesso, alla collaborazione e al vantaggio economico.[7] Finora questi benefici non sono stati ritenuti incentivi sufficienti per l'adozione generalizzata di servizi Cloud, ma si stanno sviluppando politiche, accordi contrattuali,

standard di sicurezza e procedure di controllo che potrebbero sostenere l'adozione del Cloud anche per gli archivi.

Politiche

Molti hanno chiesto una struttura internazionale coesa di politiche governative e strategie che diano indicazioni su aspetti giuridici, di sicurezza, di privacy e di condivisione dei rischi riguardanti l'ambiente Cloud. Tra questi, la Commissione Europea è la più attiva. Alla sua conferenza del 2015 sulla sicurezza nel Cloud, è stato concordato che c'è necessità sia di una politica flessibile che di approcci che permettano avanzamento tecnologico e una relazione più forte tra il settore pubblico e l'industria privata, stabilendo sicurezza in termini di network, requisiti di data location, giurisdizione straniera e accesso.[8] Tuttavia, in termini di privacy, anche se l'Europa sta sviluppando una politica unificata, sarà difficile armonizzarla con quella degli Stati Uniti, perché la privacy in Europa è considerata un diritto fondamentale ed un aspetto della dignità, mentre negli Stati Uniti è considerata un aspetto della libertà e un beneficio alienabile, a cui si può rinunciare per avere servizi personalizzati.[9]
Virginia Greiman ha condotto un'analisi comparativa delle strategie nazionali per i servizi Cloud in Australia, Unione Europea, Giappone, Singapore, Spagna, Regno Unito e Stati Uniti, e ha sviluppato alcune raccomandazioni per lo sviluppo di una strategia unificata di Cloud risultante in politiche nazionali coerenti. La prima raccomandazione è di sviluppare definizioni comuni per i termini che sono usati più comunemente, come *cyber resilience*, che generalmente si riferisce alla capacità di continuare a svolgere operazioni anche sotto attacchi, incidenti, o problemi tecnici. Queste definizioni dovrebbero essere accompagnate da tassonomie e ontologie che supportano lo sviluppo di un linguaggio comune tra il Cloud, gli Stati ed i continenti. La seconda raccomandazione è lo sviluppo di una lista uniforme di possibili problemi e delle loro cause. La terza è la creazione di una partnership tra stati che condividono gli stessi valori, come la libertà di espressione, il libero accesso all'informazione, e la protezione della privacy. La quarta è un'identificazione comune di agenti nel Cloud (per esempio, i proprietari, i fornitori di servizi, i mediatori, i trasportatori di servizi, i clienti, i controllori, e altre autorità di sorveglianza indipendenti) attraverso giurisdizioni, con indicazione delle responsabilità e dei requisiti legali. La quinta raccomandazione è la creazione di una infrastruttura per la gestione dei rischi, sviluppando norme e principi per stabilire e mantenere l'ordine civile nell'ambiente Cloud, partendo da standard internazionali. La sesta e ultima raccomandazione è un'efficace sorveglianza della sicurezza nel Cloud basata sulla trasparenza.[10]
Mentre le raccomandazioni di Greiman sono preziose e da usare come riferimento, la creazione di politiche nazionali coese e di una struttura regolamentare internazionale supportata da standards può essere aiutata da coerenti e minuziosi accordi contrattuali tra coloro che agiscono nell'ambiente Cloud.

Accordi contrattuali

Uno studio intrapreso dal progetto di ricerca InterPARES Trust[11] ha paragonato i contratti di servizi Cloud con i requisiti per la gestione, immagazzinamento e conservazione di materiali archivistici, per determinare se questi requisiti sono compresi nei contratti. I ricercatori hanno trovato che non c'è una terminologia standardizzata per il materiale archivistico considerato nei contratti per servizi Cloud, e hanno deciso di adottare il termine "dati" come la più piccola unità di informazione per far riferimento a qualsiasi tipo di materiale. Sulla base di un esame della letteratura, hanno anche trovato che i contratti per servizi Cloud comprendono parecchi documenti legali: un documento generale che delinea i servizi (per esempio i termini del servizio); un documento per ciascun servizio specifico (per esempio l'accettazione del livello del servizio); e una varietà di documenti che coprono aree come la privacy e l'uso accettabile. I requisiti per la gestione e conservazione dei dati con cui paragonare i termini dei contratti sono stati identificati dai ricercatori sulla base degli standard rilevanti ISO e ARMA, e di standard europei. Inoltre, i ricercatori hanno esaminato il documento del 2014 *Cloud Service Level Agreement Standardization Guidelines* della Commissione Europea e parecchie politiche governative. I requisiti identificati sono relativi al controllo sull'accesso, alla protezione della privacy, all' affidabilità continua e dimostrabile, all'accuratezza e autenticità dei

dati, alla trasparenza della gestione dell'account, della localizzazione di server, della distruzione dei dati e del ripristino in caso di perdita.[12]
Sulla base di quanto sopra, i contratti sono stati esaminati in relazione a questioni chiave: proprietà di dati; disponibilità, recupero e utilizzo; memorizzazione e immagazzinamento; distruzione; archiviazione e conservazione; sicurezza; localizzazione dei dati, trasferimento; e fine del servizio o conclusione del contratto.[13]

Proprietà di dati

Per quanto riguarda la proprietà di dati, il problema da considerare è che, quando un utente[14] affida i propri dati ad un fornitore o *provider* e usa l'applicazione e la piattaforma di quest'ultimo per generare dati aggiuntivi (metadati), il provider crea dei metadati relativi a queste azioni per l'elaborazione dei dati, la loro gestione, ecc. Mentre il contenuto generato e/o immagazzinato nel Cloud dall'utente è di proprietà dell'utente stesso, i metadati prodotti dal provider non lo sono, e, dato che l'utente ne ha bisogno per dimostrare l'integrità dei dati, è essenziale che i termini contrattuali determinino se e come l'utente abbia il diritto di accedere ai metadati del provider e utilizzarli.

Disponibilità, recupero ed utilizzo

Riguardo la disponibilità e l'accesso, ogni contratto dovrebbe mantenere questi due concetti legalmente distinti, perché la disponibilità è un fatto, mentre l'accesso è un diritto, ma quest'ultimo non può essere soddisfatto senza la prima. La legislazione in Nord America e in Europa garantisce il diritto alle informazioni possedute da enti pubblici e a volte anche da organizzazioni private, e queste informazioni devono essere fornite entro un determinato periodo di tempo. Quando i dati sono archiviati in un ambiente Cloud, la disponibilità dei dati archiviati implica la disponibilità delle infrastrutture, hardware e software, che facilitano il recupero e la leggibilità dei dati, perché difficoltà tecniche potrebbero rallentare il processo, e il proprietario dei dati, essendo responsabile della garanzia di accesso ai dati, potrebbe essere sanzionato.[15] Perciò gli accordi contrattuali devono specificare il grado di affidabilità del provider.
Dove l'"accessibilità" è la quantità di tempo che ci si aspetta che il sistema sia in servizio, espresso o statisticamente o in percentuale, l'"affidabilità" è la caratteristica di comportarsi in modo coerente con le aspettative.[16] Quindi, un contratto deve considerare non solo la disponibilità ma anche la "coerenza e l'accuratezza dell'accesso". Questo significa non solo che copie dei dati devono essere distribuite tra parecchi data-center, assicurando la ridondanza, ma anche che quelle copie devono rimanere coerenti quando gli utenti accedono ai dati allo stesso tempo. Questo al momento non è possibile poiché i provider non hanno espliciti accordi tra loro che aiutino ad assicurare l'accessibilità complessiva dell'Internet. Questo richiederà la collaborazione tra molteplici autorità regolamentari, come gli *stakeholder* principali, che includono i provider, gli utenti, e comunità internazionali del commercio e della standardizzazione.[17] Allo stesso tempo, i potenziali utenti dovrebbero scoprire se un provider abbia strutture che diano qualche assicurazione di affidabilità e strategie credibili di risposta se si verificasse un problema, e se il provider sia controllato da qualche autorità.

Immagazzinamento e distruzione dei dati

La questione dell'immagazzinamento e della distruzione dei dati è complessa, perché, indipendentemente da cosa sia incluso in un accordo contrattuale, la conformità è difficile da verificare. Il motivo è che l'immagazzinamento potrebbe richiedere il trasferimento dei dati da un sistema a un altro e questo potrebbe comportare la perdita dell'autenticità, mentre la distruzione potrebbe comportare la violazione di confidenzialità o della privacy, la persistenza di alcune copie e dei metadati correlati, o la persistenza di metadati generati dal provider riguardanti i dati dell'utente. I contratti standard abitualmente non contengono clausole collegate alla distruzione sistematica; al massimo assicurano che i dati degli utenti diventeranno permanentemente inaccessibili entro sei mesi, un'affermazione che non soddisfa i requisiti. Quindi, ogni accordo contrattuale deve avere degli specifici termini di servizio su a tali questioni.

Archiviazione e conservazione dei dati

L'archiviazione e la conservazione dei dati impattano la loro qualità e la loro capacità di servire come

fonti in generale e come prove legali in particolare, specialmente in cause dove l'autenticità dei dati è un'illazione basata sull'integrità del sistema in cui i dati risiedono. Gli accordi contrattuali generalmente non specificano come i dati siano mantenuti per combattere l'obsolescenza, e dicono generalmente che gli utenti sono responsabili della produzione e mantenimento di backup dei propri dati. Tutte le procedure di conservazione, compresi custodia e controllo dei dati, sono chiamate dai provider "procedure di backup".[18]

Sicurezza

La sicurezza, dal punto di vista di archiviazione e conservazione, ha a che vedere con la protezione dei dati da accessi non autorizzati, alterazione o distruzione. Il provider dovrebbe essere in grado di produrre tracce di controllo e log di accesso e di cattura, e di mantenere e rendere disponibili metadati associati agli accessi, recuperi, utilizzi e gestione dei dati, oltre ai metadati collegati ai dati stessi. I contratti standard collegano le misure di accesso ai tipi di servizi offerti e ai canoni pagati dagli utenti. Inoltre, il fatto che i dati siano stati trasferiti a un provider terzo non cambia la responsabilità del proprietario. In generale, la legge assume che la sicurezza sia esplicitamente inclusa nell'accordo contrattuale in termini tecnici, fisici e gestionali.[19] A causa di questo, la Cloud Security Alliance (CSA) sta predisponendo una direttiva sulla sicurezza come servizio, principalmente per i grandi sistemi. È logico che i grandi sistemi gestionali superino la "complessità e l'incoerenza" di molteplici piccoli sistemi, dato che forniscono economie di scala in ogni cosa, dal monitoraggio delle irregolarità all'assunzione e formazione di personale chiave, ma si può arguire che la vulnerabilità cresce con l'aggregazione di contenuto e di attività e che le preoccupazioni per la privacy sono più elevate con i grandi provider.[20]

Localizzazione e trasferimento dati

Il tema sicurezza si collega direttamente alla localizzazione dei dati e al flusso di dati oltre confine. Questo preoccupa sia in termini di leggi di protezione dati e di leggi straniere che permettono ad agenzie investigative di accedere ai dati mantenuti e registrati dai provider, sia in termini di condotta di affari regolari nella loro giurisdizione. La localizzazione dei dati può anche essere un criterio per determinare che legge da applicare nel contratto, anche se di solito i provider scelgono una giurisdizione compatibile con il loro sistema legale.[21]
Il fatto che il Cloud possa dare una localizzazione ai dati diversa dalla loro provenienza ha scatenato un dibattito sulla limitazione del movimento dei dati entro il confine del paese di provenienza, ma la strategia internazionale sta abbandonando l'idea che i dati debbano rimanere nella giurisdizione di produzione, in tal modo riconoscendo l'importanza di accordi multilaterali tra stati per una collaborazione a favore della sicurezza. Blumental si chiede se non sia presto per gli stati considerare il Cloud commerciale "infrastruttura critica" assumendo che il mercato per tali sistemi continuerà a crescere e il progresso sulla sicurezza rimarrà lento.[22]
Sicuramente, considerando che un'infrastruttura critica è costituita da cosa è importante per il funzionamento di uno stato, affidare il mantenimento e la conservazione di archivi pubblici al Cloud commerciale (public Cloud) dovrebbe sia supportare questa determinazione da parte dei governi, sia facilitare la scelta del Cloud commerciale per gli archivi correnti e storici di business e di organizzazioni non-pubbliche. Le infrastrutture critiche dipendono da altre infrastrutture, e alcuni servizi del Cloud dipendono non solo da infrastrutture elettriche e di comunicazione, ma anche da altri servizi del Cloud. C'è un potenziale per Cloud federati che si assistano a vicenda tramite la condivisione di risorse in caso di crisi e non c'è da meravigliarsi che si sia sviluppando una nuova linea di offerte di *disaster-recovery* come "servizio per emergenze" scrive Blumenthal.[23] Il fatto è che il Cloud è la piattaforma di scelta per le applicazioni per telefonini, e per i dati generati dall'utilizzo di questi e degli *smart devices* usati a casa o al lavoro, e questi dati costituiscono una percentuale crescente tra i dati nel Cloud commerciale. Quindi, è solo una questione di tempo prima che i servizi di Cloud commerciale vengano dichiarati critici, in quanto dovremo fare affidamento su di essi per mantenere e conservare i dati generati dalle piattaforme di Cloud commerciale.

Fine del servizio, conclusione di un contratto

La prospettiva di considerare il Cloud commer-

ciale un'infrastruttura critica dovrebbe limitare i timori relativi alla fine del servizio e alla chiusura del contratto. Al momento, è possibile che, se il *provider* del Cloud cessasse di esistere o terminasse uno o più dei suoi servizi (termine o sospensione dei servizi potrebbero accadere per un'infrazione, per inattività o per convenienza), i dati lasciati con il provider verrebbero cancellati o resi inaccessibili. Benché i contratti per i servizi a pagamento si esplichino per la loro durata, quelli gratuiti non hanno una durata prestabilita, e gli account possono essere chiusi unilateralmente. I contratti standard di solito richiedono che l'utente cancelli il software e le applicazioni, e potrebbero impedire all'utente l'utilizzo dei dati lasciati presso il provider. Anche quando i dati vengono restituiti all'utente, non è detto che siano in un formato utilizzabile o interoperabile. Se il contratto è concluso dall'utente, la restituzione dei dati potrebbe essere costosa ed i dati potrebbero non essere in formati accessibili. Inoltre, l'utente potrebbe non avere il diritto di accedere ai metadati generati dal sistema per archiviarli o usarli per ragioni legali, e potrebbe non avere alcuna garanzia che il provider distrugga ogni singola copia dei dati esistenti nei data-center. Quindi, la chiusura del contratto deve descrivere in modo dettagliato cosa succederà ai dati con riguardo all'accettazione dei termini di provider ed utente. Dallo studio di accordi contrattuali, e delle materie che essi trattano o devono trattare,[24] appare chiaro che l'aspetto più trascurato dai provider circa il mantenimento dei dati nel Cloud sia la conservazione. Conservare i dati nel Cloud potrebbe essere un processo da scatola nera, nella quale gli archivisti potrebbero sapere cosa inseriscono, e a cosa vogliono accedere e recuperare – verosimilmente le stesse cose che erano state inserite – ma spesso non sanno quale tecnologia è utilizzata dai provider per gestire, archiviare, o elaborare i loro dati. I provider di Cloud commerciali potrebbero anche non sapere dove fisicamente si trovino i dati, e potrebbero sotto-contrattare dei servizi ad altri provider che potenzialmente mantengono i server o sono registrati come provider in altri Paesi. Anche quando conoscono la tecnologia usata dai provider per la conservazione, gli archivisti responsabili dei dati non possono aspettarsi che lo stesso software o hardware resti in servizio per tutto il periodo necessario, o che le tecnologie che li rimpiazzano siano compatibili con quelle precedenti. Oltretutto, è improbabile che ci sia esperienza su tutte, o anche su buona parte, delle tecnologie necessarie o utilizzate per la conservazione ora e in futuro, poiché nessuno sa come le informazioni e le tecnologie di comunicazione si evolveranno.

Essendo interessati a questa questione, i ricercatori dell'InterPARES Trust hanno concluso che sia i provider che gli utenti trarrebbero benefici dallo sviluppo di uno standard internazionale sulla Preservation-as-a-Service-for-Trust (PaaST) (i.e. Conservazione come un servizio per l'affidabilità).

Preservation as a Service for Trust (PaaST)

Lo scopo dello studio chiamato Conservazione come un servizio per l'affidabilità (da qui in poi PaaST) è determinare cosa dovrebbe essere richiesto ai provider per rendere o/e mantenere affidabili i dati conservati nel Cloud; cioè perché sia possibile affidare a un provider di Cloud commerciale dati destinati alla conservazione a lungo termine, con la prospettiva di poterli recuperare identici in tutti gli aspetti essenziali, o con differenze esplicitamente identificate con tale accuratezza e precisione da essere in grado di valutare se sono adatti per ogni uso. Lo studio PaaST usa lo Unified Modeling Language (UML) che definisce un approccio tecnologico agnostico e generalizzabile per ogni tipo di dato. Il modello PaaST è sviluppato sulla base di prodotti creati nel corso dei tre precedenti progetti InterPARES e dello standard ISO Open Archival Information System (OAIS).[25] "Differisce dallo standard OAIS nel fatto che quest'ultimo è un modello di riferimento che definisce le funzioni e le informazioni necessarie per la conservazione, ma non stabilisce come potrebbero essere implementate. Sebbene PaaST sia neutrale rispetto ai metodi e alle tecnologie usate per l'implementazione, è stato sviluppato per facilitare la produzione di software capace di implementarlo. Quindi, lo scopo di PaaST è più limitato di quello dello standard OAIS, escludendo specificatamente le funzioni il cui compimento non è automatizzato, come il sollecitare e negoziare i termini di accettazione nell'Administrative Functional Entity e la produzione di raccomandazioni e piani nell'OAIS Preservation Planning Functional Entity."[26]

Il progetto PaaST è iniziato identificando i blocchi

costitutivi della conservazione, come per esempio l'applicazione di requisiti specifici di conservazione, specialmente quelli relativi alle proprietà dei dati, e ad ogni cambiamento sia dei dati che dell'hardware o del software dal quale essi dipendono, o entrambi, per mantenere l'autenticità. Questi blocchi costitutivi sono le basi dei requisiti di PaaST, che definiscono le funzionalità, i metadati, e tutte le altre informazioni necessarie per conservare i dati e produrne copie autentiche. I requisiti sono da intendere come applicabili in varie situazioni, permettendo sia l'assegnazione di diversi compiti di conservazione ad agenti differenti, sia l'esecuzione di questi compiti da parte di uno o più agenti, utilizzando diversi metodi e tecnologie. Quindi, i requisiti di conservazione devono essere articolati come servizi, cioè come serie di attività correlate che possono essere eseguite utilizzando tecnologie differenti e potenzialmente non correlate, sotto controllo amministrativo o operativo separato ed indipendente.[27] Poiché i requisiti PaaST non presuppongono o richiedono che le attività di conservazione o i controlli siano implementati in un sistema integrato, il contesto globale del processo di conservazione è chiamato *ambiente di conservazione,* piuttosto che sistema di conservazione, e definito come il complesso delle infrastrutture tecnologiche e degli strumenti utilizzati nella conservazione digitale. Tuttavia, questo non esclude la possibilità di un sistema di conservazione integrato. Inoltre, le responsabilità di conservazione possono essere distribuite in modo che alcune attività siano svolte all'interno del produttore o conservatore, mentre altre siano eseguite da uno o più provider del Cloud (per esempio, un provider potrebbe offrire archiviazione e gestione, o altri servizi specializzati, quali la migrazione dei media o la conversione di dati, in base alle necessità).
I requisiti per la conservazione sono raggruppati in serie di attività correlate chiamate *servizi*. Ogni servizio – e particolari attività all'interno di un servizio – può essere eseguito utilizzando tecnologie differenti e potenzialmente non correlate sotto un controllo operativo separato ed indipendente. I servizi di conservazione sono:

- *Submission (invio)*, che ingerisce i dati in un ambiente di conservazione;
- *Characterization (caratterizzazione)*, che identifica le proprietà tecniche, archivistiche e di rappresentazione dei dati;
- *Authenticity (autenticità)*, che cattura e riporta informazioni riguardanti l'identità e l'integrità dei dati, e l'applicazione dei metodi di autenticazione;
- *Preservation Storage (immagazzinamento)*, che controlla l'archivio dei dati per mantenerne l'identità, prevenirne la corruzione, e soddisfare gli altri requisiti di conservazione;
- *Preservation Change (cambio di conservazione)*, che governa i cambiamenti tecnologici, come la migrazione di formato o la sostituzione di software, per assicurare la sopravvivenza e l'usabilità dei dati; e
- *Access (accesso)*, che fornisce la capacità di consegnare copie di dati.

I servizi non sono necessariamente indipendenti. Per esempio, la *characterization* (caratterizzazione) sarebbe chiamata in causa dalla *submission* (invio) per decidere se accettare un set di dati inviati per la conservazione. Allo stesso modo, se richiesto da un cliente, l'*authentication* (autenticazione) potrebbe essere chiamata in causa per permettere la valutazione delle copie consegnate all'utente.
"PaaST non include servizi che non siano specificatamente appartenenti alla conservazione digitale, anche se potrebbero essere correlati, o servizi generali che ci si aspetta che un provider offra indipendentemente dal fatto che gli oggetti di informazione siano designati per una conservazione a lungo termine. I servizi generali includerebbero telecomunicazioni, management dei dati, *loggins* di sistema, sicurezza, strumenti di ricerca generica, sottosistemi di archiviazione, capacità di file transfer, etc. I servizi di conservazione utilizzeranno spesso questi servizi generali, ma la presunzione di disponibilità significa che essi non devono essere articolati all'interno dei requisiti PaaST. Un esempio comune di servizio correlato e che non è incluso in PaaST è il sistema di holding del management spesso utilizzato dalle istituzioni come archivi e biblioteche per gestire gli inventari dei materiali dei quali loro sono responsabili."[28]
Coloro che gestiscono i servizi sono individuati come quattro ruoli primari: l'*Initial Holder,* che mantiene, possiede o controlla i dati da conservare, il *Preservation Director,* che ha responsabilità per la preservazione dei dati; il *Preservation Service Provider,* che fornisce le risorse tecnologiche e i servizi; e l'*Access Client,* che richiede l'accesso ai dati conservati. C'è un ruolo secondario di *Submitter,* che materialmente invia i dati al Provider ed è respon-

sabile per rispondere alle domande del Provider.
I requisiti funzionali per ogni servizio sono numerati in maniera unica. Delle pre-condizioni devono esistere prima che il servizio possa operare. Un flusso principale dettaglia le operazioni sequenziali del servizio, e un flusso alternativo rimarca gli errori e le condizioni di eccezione nel controllo che potrebbero sorgere. Il PaaST è supportato da specifici termini e condizioni per la conservazione che sono raggruppati in due documenti: l'accordo di conservazione e un contratto di servizio di conservazione. Il diagramma UML è accompagnato da studi di casi. Le operazioni e gli attributi seguono una convenzione standardizzata di denominazione che supporta l'accesso e la riusabilità attraverso altre operazioni e servizi.
L'obiettivo del progetto InterPARES Trust (o ITrust) è di generare i quadri teorici e metodologici necessari per lo sviluppo di politiche, procedure, regolamentazioni, standard, legislazioni locali, nazionali ed internazionali, che siano in grado di assicurare la fiducia pubblica nei documenti digitali. La ricerca ha prodotto risultati che consentono l'adozione di un sistema di conservazione capace di esistere in uno o più Cloud e in un "ambiente ibrido". Questa visione è riflessa nel design di un ambiente di conservazione affidabile attraverso la definizione di requisiti, attività, attori e competenze, entità e relazioni, usando un modello UML che ha il potenziale di implementazione in una gran varietà di contesti.

Conclusione

Come saranno in futuro i sistemi di conservazione? Probabilmente non assomiglieranno a sistemi nel senso tecnologico del termine. Piuttosto, saranno costituiti di parti connesse che produrranno un complesso globale che si spera sia governato da un "quadro" comune di principi, regole e procedure. Saranno ibridi – comprendendo servizi nel Cloud e *in-house*, ma la loro capacità di rimanere coesi, integrati, indipendenti, interoperabili, flessibili, accessibili ed affidabili dipenderà dai futuri sviluppi tecnologici e dai benefici/vantaggi economici che saranno capaci di offrire.
L'affidabilità della gestione e conservazione degli archivi in questi sistemi dipenderà direttamente dall'affidabilità dei servizi del provider, dalla sicurezza dell'architettura dell'infrastruttura del Cloud, e dalle sue procedure. Se questi sistemi verranno considerati efficaci per la conservazione permanente dipenderà dalla capacità dei provider del Cloud di interconnetterli verticalmente (tramite provider di servizi specializzati che si appoggiano su provider più grandi) e orizzontalmente, con una "federazione" che offra non solo ridondanza, ma accesso universale e tutti i tipi di scambio. Per quanto riguarda l'autenticità del materiale, ci affideremo a questi sistemi augurandoci che la continuità della ricerca internazionale ed interdisciplinare sarà in grado di garantire che essi rimangano un interesse centrale degli stati nei loro accordi.

NOTE

[1] William Lehr, *Reliability and the Internet Cloud*, in *Regulating the Cloud. Policy for Computing Infrastructure*, a cura di Christopher S. Yoo e Jean-François Blanchette, Cambridge, Massachusetts and London, England: The MIT Press, 2015, p. 336-350.
[2] Jean-François Blanchette, *Introduction*, in *Regulating the Cloud. Policy for Computing Infrastructure*, cit., p.3.
[3] Ibidem, p. 5.
[4] Joe Weinman, *Cloud Strategy and Economics*, in *Regulating the Cloud. Policy for Computing Infrastructure,* cit., p. 25-28 e p. 37-38.
[5] NIST Cloud Computing Standards Roadmap Working Group, *NIST Cloud Computing Standards Roadmap*, NIST Special Publication 500-291, version 2. US Department of Commerce, National Institute of Standards and Technology, July 2013.
[6] Luciana Duranti, *Preservation in the Cloud: Towards an International Framework for a Balance of Trust and Trustworthiness*, in *APA/C-DAC International Conference on Digital Preservation and Development of Trusted Digital Repositories. 5-6 February 2014. New Delhi, India,* a cura di Dinesh Katre e David Giaretta, New Delhi: Excel India Publishers, 2014, p. 23-38.
[7] Cfr. Luciana Duranti, *Archival Science in the Cloud Environment: Continuity or Transformation?*, in *Atlanti* vol. 23(2013), p. 45-52; Luciana Duranti - Corinne Rogers, *Trust in digital records: An increasingly cloudy legal area*, in *Computer Law & Security Review* 28.5, October 2012, p. 522-531; Luciana Duranti - Corinne Rogers, *Trust in online records and data*, in *Integrity in Government through Records Management: Essays in Honour of Anne Thurston*, a cura di James Lowry e Justus Wamukoya, Farnham: Ashgate, 2014, p. 203-216; Luciana Duranti, *Digital Records and Archives in the Commercial Cloud*, in *Regulating the Cloud. Policy for Computing Infrastructure*, cit., p. 197-214. Vedi anche *The Canadian Journal of Information and Library Science*, special issue on Data, Records, and Archives in the Cloud, guest editor Luciana Duranti. Vol. 39, n. 2, June 2015.
[8] The European Network and Information Security Agency (ENISA), *Security Framework for Government Clouds,* February 2015.

[9] Andrea Renda, *Cloud Privacy Law in the United States and the European Union*, *Regulating the Cloud. Policy for Computing Infrastructure*, cit., p. 135-164.

[10] Virginia Greiman, *National Strategies for Cloud Innovation and Security*, in *Proceedings of the 3rd International Conference on Cloud Security Management. University of Washington – Tacoma, USA 22-23 October 2015*, a cura di Barbara Endicott-Popovski, Reading, UK: ACPI, 2015, p. 46-57.

[11] InterPARES Trust (ITrust 2013-2018 – www.interparestrust.org) è un progetto di ricerca finanziato da un grant del Social Sciences and Humanities Research Council of Canada. ITrust esplora i problemi relativi ai documenti digitali affidati all' Internet. ITrust è la quarta fase del progetto International Research on Permanent Authentic Records in Electronic Systems (InterPARES – 1998-2018, www.interpares.org).

[12] Jessica Bushey - Marie Demoulin - Robert McLelland, *Cloud Service Contracts: An Issue of Trust*, in *The Canadian Journal of Information and Library Science,* Vol. 39, no. 2, June 2015, p. 128-153. Vedi anche la *Checklist for Cloud Service Contracts* sul sito pubblico di InterPARES Trust: https://interparestrust.org/assets/public/dissemination/NA14_20160226_CloudServiceProviderContracts_Checklist_Final.pdf.

[13] Ibidem, p. 135.

[14] Nel contesto di questa relazione, il termine utente si riferisce a chiunque usi i servizi di un Cloud provider.

[15] Jessica Bushey - Marie Demoulin - Robert McLelland, *Cloud Service Contracts: An Issue of Trust,* cit., p. 137-138.

[16] William Lehr, *Reliability and the Internet Cloud*, in *Regulating the Cloud. Policy for Computing Infrastructure*, cit., p. 95.

[17] Ibidem, pp. 100-101.

[18] Jessica Bushey - Marie Demoulin - Robert McLelland, *Cloud Service Contracts: An Issue of Trust*, cit., p.140.

[19] Ibidem, p. 141.

[20] Marjorie Blumenthal, *Finding Security in the Cloud*, in *Regulating the Cloud. Policy for Computing Infrastructure*, cit., p. 64.

[21] Elaine Goh, *Clear skies or cloudy forecast? Legal challenges in the management and acquisition of audiovisual materials in the cloud*, in *Records Management Journal*, Vol. 24, n.1, 2014, p.59.

[22] Marjorie Blumenthal, *Finding Security in the Cloud*, cit., p. 65.

[23] Ibidem, p. 68.

[24] Questo studio ha prodotto una lista di materie che devono essere incluse in contratti tra Cloud provider e user. La lista è accessibile qui: https://interparestrust.org/trust/research_dissemination, sotto InterPARES Trust Research Documents, NA14.

[25] International Standards Organization. Space data and information transfer systems – Open archival information system (OAIS) – reference model. ISO 14721:2012. http://www.iso.org/iso/catalogue_detail.htm?csnumber=57284

[26] Luciana Duranti - Adam Jansen - Giovanni Michetti - Courtney Mumma - Daryll Prescott - Corinne Rogers - Kenneth Thibodeau, *Preservation as a Service for Trust (PaaST)*, in *Security in the Private Cloud*, a cura di John R. Vacca, CRC Press - an imprint of Taylor & Francis Group, LLC, in corso di stampa.

[27] Ibidem.

[28] Ibidem.

DOI: 10.3302/0392-8586-201606-057-1

ABSTRACT

Several countries are beginning to look at the Internet Cloud as a critical meta-infrastructure that is vital to the functioning of their economy and society. This article contends that, in the future, recordkeeping and preservation systems will be more often than not in the cloud; identifies issues related to contractual agreements; and presents research carried out in the context of the InterPARES Trust project about the development of Preservation as a Service for Trust (PaaST). It concludes that whether these services will be "trustworthy" will depend on the ability of records professionals to develop standards for an international framework for data and records in the cloud, and on their impact on government policy and the public opinion.