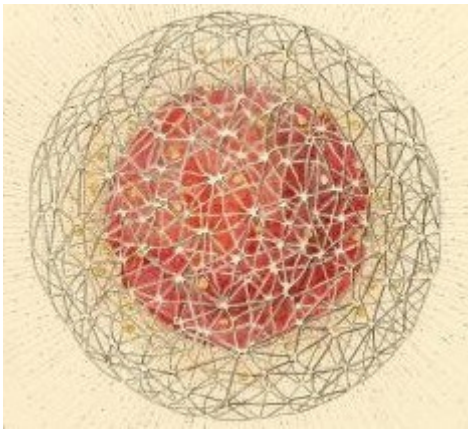
Associazione Nazionale Archivistica Italiana–Direzione Generale per gli Archivi



Il Mondo degli Archivi – STUDI

A. II - Ottobre 2014

Gli archivi come Data Center

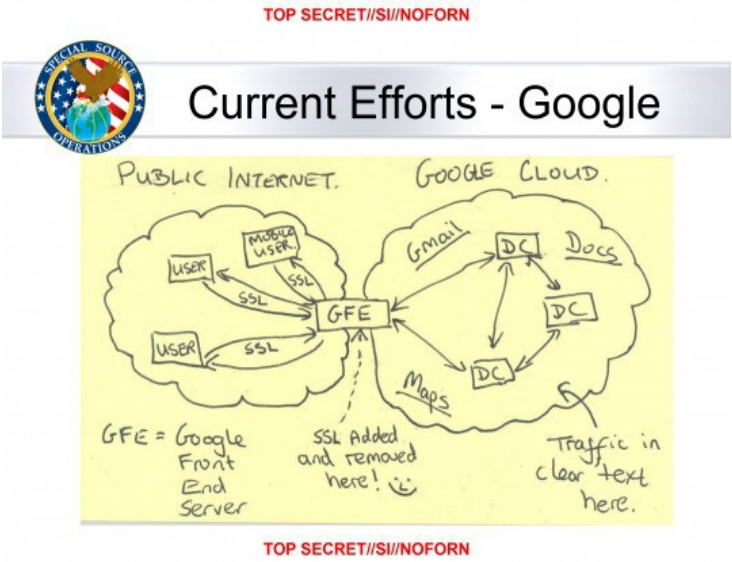
di Simone Vettore

Registrazione presso il Tribunale di Roma n. 527/93 del 29.11.1993 – ISSN 2284-0494

Il saggio è pubblicato al seguente indirizzo: http://www.ilmondodegliarchivi.org/index.php/studi/item/439-gli-archivi-come-data-center

***

Nel pieno dello scandalo *datagate* fece particolarmente scalpore la pubblicazione, da parte del Washington Post, di una *slide* preparata dagli agenti della National Security Agency nella quale si illustrava, con disegni decisamente *naif* e con tanto di faccina sorridente a suggellare il trionfo[1], lo stratagemma adottato da questa agenzia per "intercettare" le comunicazioni di ignari cittadini.



In sostanza l'NSA riusciva ad intercettare i dati insinuandosi nel delicato momento di passaggio tra l'Internet pubblica e quella privata, proprio là dove viene attivato e rimosso il protocollo crittografico SSL che, in linea teorica, dovrebbe garantire la sicurezza della trasmissione dei dati.

Diversamente dalla maggior parte degli addetti ai lavori e dei commentatori, chi vi scrive non è rimasto particolarmente scandalizzato, né tantomeno sorpreso, dall'apprendere che questa agenzia federale operasse ben al di là dei paletti imposti dal FISA (Foreign Intelligence Surveillance Act): in un ambiente per definizione *borderline* qual è quello dei "servizi segreti" ci vuole infatti una buona dose d'ingenuità a credere che si rispettino pedissequamente le regole!

Purtroppo nelle successive analisi si è continuato ad appuntare l'attenzione soprattutto sull'aspetto "formale" della vicenda (vale a dire la violazione delle leggi da parte dell'NSA, fatto naturalmente esecrabile) od al più si è cercato di delineare l'esatta natura dei rapporti tra governi ed aziende IT.

Ben pochi però si sono spinti, con la loro analisi, al livello ulteriore e che rappresenta, è opinione di chi scrive, il vero nocciolo della questione: il ruolo vitale svolto nella società contemporanea da quei *data center* (DC) che, proprio in quanto "archivi" delle nostre esistenze digitali, sono risultati essere uno dei principali "obiettivi" delle intercettazioni.

[1] Particolare che, ricordano gli autori dello *scoop* sul WP, sommato allo sfrontato titolo della *slide*, ovvero "Google Cloud Exploitation", fece andare su tutte le furie i tecnici dell'azienda di Mountain View. Vedi B. Gellman – A. Soltani, *NSA infiltrates links to Yahoo, Google data centers worldwide, Snowden documents say*, http://www.washingtonpost.com/world/national-security/nsa-infiltrates-links-to-yahoo-google-data-centers-worldwide-snowden-documents-say/2013/10/30/e51d661e-4166-11e3-8b74-d89d714ca4dd_story.html.

Come si spiega, fatta salva qualche rara eccezione, questo generale disinteresse da parte dei *media mainstream*? Probabilmente, essendo la materia eccessivamente tecnologica e difficilmente "digeribile" da parte del grande pubblico, si è preferito puntare su argomenti più "vendibili": vuoi mettere l'impatto mediatico e, di conseguenza, le vendite dal giornalaio tra il descrivere il funzionamento di un DC e la reazione irata della Merkel che chiama Obama?
In questo breve saggio, pertanto, cercheremo di colmare questa lacuna ponendo al centro della nostra analisi proprio gli archivi / *data center*, spiegando cosa sono, come funzionano e soprattutto perché sono così importanti al punto da svolgere, pur su differenti piani, una importante funzione "politica".

Partiamo dunque dall'inizio: cos'è un *data center*? Sostanzialmente si tratta di un luogo fisico attrezzato con tutti gli impianti necessari per ospitare e far funzionare un numero variabile di elaboratori elettronici posti al servizio di una o più organizzazioni.
In questa definizione ritroviamo evidenziati, seppur talvolta in modo solo implicito, tutti quegli elementi / caratteristiche che sono proprie di un DC a prescindere dalle sue dimensioni, vale a dire, scendendo maggiormente nel dettaglio:

1. la sua natura fisica, a dispetto di un paradigma tecnologico dominante che lo vorrebbe quasi etereo e "tra le nuvole"
2. tutti quegli impianti propedeutici al funzionamento del DC stesso, quali sistemi di alimentazione, di condizionamento, etc.
3. gli elaboratori deputati ad effettuare le operazioni di calcolo affiancati dai non meno imprescindibili sistemi di *storage* / archiviazione e di telecomunicazione.

Se questa è la teoria, nella pratica i *data center* sono tipicamente contraddistinti dalla presenza, all'interno di un medesimo spazio fisico (di norma un grande capannone), di migliaia di elaboratori di fascia medio – bassa interconnessi tra di loro; Google, che è stata tra le prime aziende ad applicare tale modello, lo definisce come *wharehouse computing* e lo descrive nei seguenti termini:

The hardware for such a platform consists of thousands of individual computing nodes with their corresponding networking and storage subsystems, power distribution and conditioning equipment and extensive cooling systems. The enclosure for these systems is in fact a building structure and often indistinguishable from a large warehouse[2].

[2] Vedi L. A. Barroso – U. Hölzle, *The data center as a Comupter. An Introduction to the Design of Warehouse-scale Machines*, s.l., Morgan & Claypool, 200, p. 2, scaricabile al seguente URL: http://www.morganclaypool.com/doi/pdf/10.2200/S00193ED1V01Y200905CAC006. Un *data center* concepito secondo questi criteri può alternativamente essere chiamato come *server farm* ("fattoria di server", giusto a sottolineare la massiccia presenza di questi ultimi) o

Da notare, nel passaggio appena citato, la particolare enfasi posta dagli ingegneri di Google su quel complesso di sottosistemi che ricoprono un'importanza cruciale nell'assicurare la sicurezza e la piena operatività dei *data center*.

E non a caso questi ultimi sono alcuni degli argomenti maggiormente trattati dalla letteratura specialistica, la quale distingue tra misure volte ad assicurare la sicurezza fisica e quelle tese a garantire la sicurezza logica.

La prima la si raggiunge:

1. realizzando i *data center* in siti non esposti a rischio sismico, idro-geologico, etc.
2. ponendo in essere rigidi sistemi di sorveglianza, di controllo degli accessi, etc.
3. implementando una serie di sistemi tra i quali vanno assolutamente menzionati i seguenti:
   - UPS (Uninterruptible Power Supply), il quale a) garantisce l'erogazione continua di energia elettrica alla struttura e, qualora dovesse verificarsi un'interruzione nella fornitura da parte della *public utility*, b) fa intervenire la batteria fintantoché non interviene il generatore di emergenza, il tutto c) senza che si verifichino sbalzi di tensione dannosi per le "macchine"
   - PDU (Power Distribution Units), ovvero il complesso sistema di distribuzione dell'energia elettrica attraverso quadri e/o interruttori elettrici solitamente "annegati" nel pavimento del *data center*
   - sistema di condizionamento, che in genere opera attraverso CRAC (Computer Room Air Conditioning), vale a dire "stanze" dalle quali spira aria fredda che, fluendo sotto il pavimento, esce attraverso delle grate giusto in corrispondenza dei *rack* da raffreddare.

La seconda invece richiede l'adozione dei più rigidi protocolli in ordine alla sicurezza dei dati (AES, TLS, etc.) nonché di stringenti procedure e *policy* interne, sovente ispirate da / basate su appositi standard internazionali.

Nonostante tutti questi accorgimenti, sono stati proprio DC come quelli appena descritti ad essere oggetto delle "attenzioni" dell'NSA ed il motivo va rintracciato, fondamentalmente, per il ruolo centrale da essi svolto all'interno dell'odierno panorama tecnologico.

Quest'ultimo è caratterizzato, come sarà noto, dal modello del *cloud computing* e da quello, complementare, del *fog computing*. Con il primo si intende quell'insieme variegato di servizi, *software* ed infrastrutture *hardware* offerti da un *service provider* ed accessibili via Internet da un qualsiasi dispositivo[3]; al di là della definizione forse un po' astrusa, si tratta di un paradigma con il

webfarm (quest'ultimo però con una più spiccata propensione a fornire servizi Internet).

[3] A tal proposito è utile ricordare l'ulteriore suddivisione tipologica dei servizi erogati in modalità *cloud*: a) SaaS (Software as a Service) b) PaaS (Platform as a Service) c) Iaas (Infrastructure as a Service). A questa triade se ne potrebbero aggiungere altre più "archivistiche" come DaaS (Data as a Service), EaaS (Emulation as a Service), etc.

quale tutti noi abbiamo a che vedere nel nostro quotidiano popolato di reti sociali, *app* e servizi *online* di ogni genere ai quali accediamo da remoto attraverso una panoplia di dispositivi fissi e mobili (quali *smartphone*, *tablet*, etc.).
Il concetto di *fog computing*, teorizzato dagli ingegneri di Cisco, non si discosta di molto dal precedente (l'obiettivo è ugualmente l'erogazione di servizi da remoto) ma ambisce a sviluppare un'infrastruttura geograficamente più diffusa nonché più vicina agli utenti finali in modo tale, tra l'altro, da risultare adatta a quelle che sono considerate le nuove frontiere del *computing*, vale a dire l'Internet delle Cose ed il *wearable computing* [4].
Al fine di comprendere affinità e diversità tra questi due modelli è opportuno riportare le esatte parole dei tecnici di Cisco:

Fog Computing is a paradigm that extends Cloud computing and services to the edge of the network. Similar to Cloud, Fog provides data, compute, storage, and application services to end-users. The distinguishing Fog characteristics are its proximity to end-users, its dense geographical distribution, and its support for mobility. Services are hosted at the network edge or even end devices such as set-top-boxes or access points. By doing so, Fog reduces service latency, and improves QoS, resulting in superior user-experience. Fog Computing supports emerging Internet of Everything (IoE) applications that demand real-time/predictable latency[5].

Da questo breve passaggio si desume che, se la tipologia di servizi offerti è pressoché sovrapponibile, a far la differenza è l'infrastruttura, vale a dire i *data center*, che li supporta; se nel caso del *cloud computing*, come visto, siamo di fronte a costruzioni *monstre*, nel caso del *fog computing*, benché non vengano specificate la caratteristiche, si può desumere si tratti di DC di più piccole dimensioni ma costruiti in maggior numero e, soprattutto, dotati di "macchine" più performanti. Solo in tal modo, infatti, possono venir soddisfatti i requisiti richiesti di ampia distribuzione geografica e di ridotti tempi di latenza!

Sorvolando ora sulle implicazioni di questa differenziazione circa le modalità di realizzazione dei DC (sulle quali comunque torneremo più avanti), è qui il caso di concentrarsi su quello che è il passaggio centrale di questa analisi: i *data center*, siano essi al servizio della "nuvola informatica" o della "nebbia", oltre che per le capacità di calcolo che offrono a servizi che lavorano sempre più dal lato *server*, sono di importanza fondamentale per i dati ed i documenti che essi contengono: foto,

[4] Con l'Internet delle Cose (*Internet of Things*, o IoT) si intende l'estensione e la connessione ad Internet degli oggetti e dei luoghi (dagli elettrodomestici alle auto, dalle case agli uffici) mentre con il *wearable computing* si individua tutta quelle serie di dispositivi tecnologici indossabili (come ad es. *smartwatch*, braccialetti intelligenti, *smart glasses*, etc.).
[5] Vedi *Fog Computing, Ecosystem, Architecture and Application*, http://www.cisco.com/web/about/ac50/ac207/crc_new/university/RFP/rfp13078.html.

messaggi, *e-mail*, video, *file* di ogni tipo, appartenenti a semplici cittadini così come ad organizzazioni pubbliche e private, si vanno accumulando al loro interno al punto che essi si vanno configurando sempre più come gli archivi digitali per antonomasia.
Appare dunque evidente l'interesse che essi suscitano non solo per le varie agenzie come l'NSA ma anche per *cyber terroristi*, pirati informatici e malintenzionati vari: poter mettere le mani sui dati / documenti[6] contenuti in un *data center* significa accedere ad una mole pressoché sterminata di informazioni. Informazioni che, per inciso, essendo **le nostre** dovrebbero portare il tema al centro del dibattito pubblico, cosa che purtroppo non avviene stante la complessiva scarsa informazione.

Alla luce di quanto fin qui scritto risulta chiaro perché, nella *slide* presentata all'inizio, tutta l'attenzione dell'NSA fosse incentrata nello "sfruttare" al massimo i *server* che compongono la *cloud* di Google: resta solo da spiegare, compatibilmente con quanto filtrato sui media, come nel concreto avvenisse l'*exploitation*[7].
Preliminarmente però occorre spiegare cosa accade quando inviamo una *e-mail* con GMail, carichiamo un *file* su Drive (il servizio di *cloud storage* di Google, n.d.r.) o comunque adoperiamo uno dei numerosi servizi di BigG; in estrema sintesi quando compiamo una di queste azioni parte, sull'Internet pubblica, una comunicazione protetta tra il *device* che stiamo adoperando ed uno dei vari Google Front End Server; quest'ultimo funge da punto di raccordo / smistamento delle istanze tra l'Internet pubblica e la rete privata di Google, all'interno della quale le comunicazioni, fino alle rivelazioni di Snowden, avvenivano in chiaro in quanto ritenuta sufficientemente sicura. I GFE, a monte dei quali verosimilmente si trova un Network Load Balancer con il compito di ripartire i "carichi di lavoro", appena ricevuta l'istanza la inoltrano ai vari *server* interni sui quali risiedono i dati e, una volta ottenuta risposta, la rispediscono al richiedente.
Un particolare balza agli occhi: i nostri dati / documenti digitali, contrariamente a quanto potremmo pensare, non solo si trovano in un unico data center ma nemmeno mantengono la loro unità "fisica". Si tratta di una precisa politica di Google la quale li spezzetta, li replica e li fa circolare in continuazione tra i vari DC sparsi sul globo al fine di aumentarne le probabilità di sopravvivenza. Spiega l'azienda californiana a riguardo:

Anziché memorizzare i dati di ogni utente in una singola macchina o insieme di macchine, tutti i dati (compresi i nostri) vengono distribuiti tra più computer in luoghi diversi. In seguito, tali dati vengono

[6] Ma anche, come spesso accade, ai soli metadati, stante la possibilità (specie quando posseduti in gran quantità) di poterli incrociare e rielaborare, ricavandone ugualmente utili informazioni.
[7] Essendo plausibile che, magari con qualche leggero aggiustamento, un simile *modus operandi* venisse seguito anche nei confronti dei *network* di altre società, prendiamo Google come modello esplicativo di riferimento.

suddivisi e replicati in più sistemi per evitare single point of failure. Come ulteriore misura di sicurezza, a tali porzioni di dati viene assegnato un nome casuale, che li rende illeggibili all'occhio umano[8].

In una siffatta architettura i punti deboli erano, con tutta evidenza, almeno due: 1) i GFE, ossia là dove si passa da una connessione crittografata ad una in chiaro 2) l'insieme delle comunicazioni in chiaro all'interno della rete privata[9].

Stando alle ricostruzioni giornalistiche l'NSA avrebbe sfruttato la prima delle due vulnerabilità anche se, al riguardo, non è ancora stato reso noto con precisione se l'intercettazione sia avvenuta violando i GFE o se l'NSA sia stata in grado di decifrare il protocollo SSL[10] e, dunque, di "insinuarsi" prima che la comunicazione (= il pacchetto di *bit*) entrasse nella *private network*.

Abbastanza "quotata" resta comunque anche l'ipotesi che l'NSA sia stata in grado di penetrare *tout court* all'interno della rete privata di proprietà di Google e, una volta dentro, di intercettare le comunicazioni tra un *data center* e l'altro; inutile aggiungere che se tale scenario (corroborato, a detta di molti, dalla ricordata corsa alla criptazione delle comunicazioni interne, cfr. n. 9) si fosse realmente concretizzato si tratterebbe di una disfatta completa di tutte le procedure di sicurezza fin qui messe in atto.

Accanto alla spiegazioni appena fornite, che paiono essere le più probabili, ne sono circolate di ulteriori che, pur risultando essere più che altro delle "varianti sul tema", è comunque opportuno riportare: la prima ipotesi alternativa è che l'NSA sia riuscita ad introdursi in qualche Internet Exchange Point, una sorta di punto di raccordo nel quale, attraverso una serie di *switch* e di *router*, vengono interconnesse reti appartenenti ad Internet Service Provider diversi (Google, possedendo una propria rete, è *de facto* anche un ISP, n.d.r.). La seconda, concettualmente affine alla precedente, è che l'intromissione nella *private network* sia avvenuta a livello di *cable landing station*, ovvero quel punto (precisamente individuabile) nel quale i cavi sottomarini in fibra ottica riemergono dalle profondità marine[11]. La terza ed ultima ipotesi è che l'NSA si sia intrufolata all'interno di *data center* che Google condivide con altri operatori.

---

[8] Vedi http://www.google.com/about/datacenters/inside/data-security/index.html.

[9] La scelta di trasmettere in chiaro derivava dalla volontà di non rallentare il processo con operazioni di criptazione / decriptazione; dopo le rivelazioni di Snowden si è corso ai ripari, provvedendo alla progressiva criptazione di tutte le comunicazioni interne, evitando naturalmente che ciò andasse a detrimento delle prestazioni. Vedi D.A. Sanger – N. Perlroth, *Internet Giants Erect Barriers to Spy Agencies*, http://www.nytimes.com/_2014/06/07/technology/internet-giants-erect-barriers-to-spy-agencies.html.

[10] Vedi Steven J. Vaughan-Nichols, *Google, the NSA and the need for locking down datacenter traffic*, http://www.zdnet.com/google-the-nsa-and-the-need-for-locking-down-datacenter-traffic-7000022632/ e, dello stesso a., *Has the NSA broken SSL? TLS? AES?*, http://www.zdnet.com/has-the-nsa-broken-ssl-tls-aes-7000020312/.

[11] Sull'argomento ci si permette di rinviare al proprio *Sistema di comunicazioni globale e relazioni internazionali*, http://www.bloglobal.net/2013/11/sistema-di-comunicazioni-globale-e-relazioni-internazionali.html.

Sedi dei data center

Gestiamo e siamo proprietari di data center in tutto il mondo per garantire il funzionamento dei nostri prodotti 24 ore su 24, 7 giorni su 7. Scopri ulteriori informazioni sulle sedi dei nostri data center, sul nostro impegno presso le comunità locali e sulle opportunità di lavoro in tali sedi.

Americhe

Contea di Berkeley, Carolina del Sud

Council Bluffs, Iowa

Contea di Douglas, Georgia

Quilicura, Cile

Contea di Mayes, Oklahoma

Lenoir, Carolina del Nord

The Dalles, Oregon

Asia

Hong Kong

Singapore

Taiwan

Europa

Hamina, Finlandia

Saint-Ghislain, Belgio

Dublino, Irlanda

Sia come sia resta il fatto, grave, che il sistema nel suo complesso presenta delle falle di sicurezza e che sono in molti a volerle sfruttare; al riguardo è opinione di chi scrive che una chiave di lettura feconda per valutare le conseguenze - e formulare possibili risposte – sia proprio quella "archivistica", la quale discende a sua volta dalla convinzione, già espressa, che i *data center* siano a tutti gli effetti gli archivi (digitali) contemporanei.

I livelli di analisi possibili sono molteplici ma nel contempo intimamente interconnessi tra di loro: in estrema sintesi al livello più basso, che potremmo definire "documentario", si situano i problemi collegati con la produzione, sedimentazione e conservazione di dati e documenti mentre al livello intermedio troviamo il ruolo che lo Stato deve (o meno) svolgere in materia (politica archivistica); strettamente legata a quest'ultima funzione, ma su un piano ancora più elevato, la natura stessa dello Stato ed i rapporti che quest'ultimo intrattiene con gli altri attori attivi nell'agone delle relazioni internazionali.

Cerchiamo dunque di scendere un po' più nel dettaglio, chiedendo venia anticipatamente se l'esposizione non potrà essere perfettamente lineare.

Già al primo livello le questioni sollevate dallo scandalo *datagate* sono alquanto spinose: se nel caso di intercettazione mediante violazione del protocollo SSL si può parlare, in termini archivistici, di corruzione dei singolo dato / documento, le conseguenze non sono in fin dei conti così catastrofiche giacché potrebbe darsi che quest'ultimo, in forza delle citate operazioni di allineamento / retroallineamento, possa venir "recuperato"; ben diverso il caso di violazione *tout court* del DC (come purtroppo potrebbe esservi verificato) in quanto in tale evenienza a venir meno è l'affidabilità complessiva del sistema e, con essa, il requisito imprescindibile dell'ininterrotta custodia (*unbroken custody*) dell'intera massa dei documenti conservati.
Relativamente a quest'ultimo principio è doverosa una digressione: mentre in ambiente analogico esso si basa sull'assunto che la qualità archivistica dipende dalla possibilità di provare l'esistenza di una serie ininterrotta di custodi responsabili dei documenti, in ambiente digitale l'autenticità di un documento è ottenibile a patto di poter dimostrare, attraverso *audit trail* e rigorose procedure di controllo, che dalla creazione del dato / documento (con contestuale inserimento nel sistema di gestione) fino al suo trasferimento nel sistema di conservazione di *long-term*, esso sia stato sempre al sicuro all'interno di detti sistemi[12] e che in nessun caso essi siano stati violati.

Il principio dell'*unbroken custody* poc'anzi ricordato ed il ruolo imprescindibile in esso svolto dal "custode" ci introduce al secondo livello in quanto, in particolar modo nei paesi di diritto romano, lo Stato è da sempre considerato il custode per eccellenza (e non solo dei propri documenti ma anche di quelli prodotti da soggetti terzi!); anzi, la presenza di un archivio, in quanto testimonianza dell'azione statuale, ha rappresentato e rappresenta, anche dal punto di vista simbolico, manifestazione della sua sovranità ed attestazione della sua stessa esistenza.
Non sfuggono dunque, viste da questa particolare prospettiva, le implicazioni profonde delle trasformazioni descritte nelle pagine precedenti: il fatto che gli archivi contemporanei coincidano sempre più con i grandi *data center* e che questi ultimi appartengono ai colossi dell'informatica non rappresenta forse l'ulteriore prova della progressiva abdicazione dello stato nazionale di struttura gerarchico-piramidale, così come l'abbiamo ereditato dalla Rivoluzione francese, in favore di nuovi soggetti transnazionali e "liquidi"?
A nostro parere, evidentemente, la risposta a tale quesito non può che essere affermativa anche se, a costo di ripeterci, è doveroso ricordare la presenza di aspetti paradossali che offuscano un quadro altrimenti cristallino.

[12] Sistema di gestione e sistema di conservazione di lungo periodo, seppur logicamente distinti sono, all'atto pratico, sempre più coincidenti agli occhi di operatori ed utenti. In questo saggio, si sarà intuito, assumiamo che gran parte del *life-cycle* di dati e documenti avvenga all'interno di sistemi che a loro volta "vivono" all'interno dei *data center*.

Il comportamento tenuto dal Governo degli Stati Uniti è, ancora una volta, emblematico: se da un lato esso risulta perdere terreno nei confronti di (alcuni) colossi dell'ICT, al punto da sentire l'esigenza di spiarne i contenuti, dall'altra ha dimostrato di possedere capacità tecnologiche non indifferenti, riuscendo a violare i sistemi di difesa da costoro approntati. A questo punto resta da chiedersi come mai esso non provveda a costruirsi DC in proprio, continuando in tal modo ad esercitare la funzione archiviale che gli è propria, ma al contrario finisca per avvalersi (facendosi *client*) pure esso dei servizi di *cloud storage* forniti da *provider* presenti sul mercato[13]!

Le possibili risposte, che spaziano dalle esigenze di contenimento della spesa pubblica alla motivazione tutta ideologica di lasciar spazio all'iniziativa privata, non paiono del resto soddisfacenti; piuttosto crediamo che si possa attribuire tale comportamento schizofrenico alle esigenze contingenti di questo particolare frangente storico nel quale l'istituzione statuale è sì in declino ma controlla ancora importanti leve di potere: l'amministrazione statunitense, avvalendosi dei rapporti spesso privilegiati esistenti tra vertici militari e *top management* dei colossi della Silicon Valley, non è, nonostante tutto, stata in grado di ottenere la "predisposizione" di *backdoor* (e talvolta di *frontdoor*) che consentono l'accesso ai sistemi che si intende "attenzionare" e, per di più, di affiancare a questi metodi *soft* pratiche ben più invasive?
Simili dinamiche, del resto, sono individuabili pressoché in tutti gli stati: in Italia, ad esempio, già lo scandalo Telecom – SISMI del 2006 aveva portato alla luce la rete di favori e connivenze esistente tra servizi segreti e vertici dell'ex monopolista delle TLC, il cui valore strategico per la sicurezza nazionale (al di là delle violazioni commesse nel caso specifico) veniva in tal modo ribadito[14]. Non diversamente vanno le cose in Cina dove, in linea con il capitalismo di stato, l'esistenza di rapporti speciali tra aziende come Lenovo, Huaweii, ZTE, etc. ed il governo di Pechino sono dati per assodati. In Russia, per finire questa rapida carrellata, nel corso dell'ultimo anno abbiamo assistito ad una stretta del controllo statale su Internet e sulle aziende che vi operano: a marzo si è consumato il "caso" Vkontakte (noto in Occidente come il "Facebook russo"), con il fondatore Pavel Durov costretto a lasciare il proprio posto di amministratore delegato ad Igor Sechin[15]; nelle ultime settimane, infine, sulla scia dell'inasprirsi delle relazioni con l'Occidente a

[13] Al riguardo è spiazzante scoprire che si richiede ai vari *player* di realizzare, adottando requisiti assai stringenti, apposite G-Cloud (*Governative Cloud*) e che nel contempo si pongono le stesse aziende fornitrici sotto la lente d'ingrandimento dell'*intelligence*.

[14] Telecom Italia, per inciso, è attiva nel campo dei servizi *cloud* con prodotti come Nuvola Italiana e TI Cloud.

[15] Il Financial Times definisce Sechin, che è pure presidente di Rosneft (la compagnia petrolifera statale russa) e naturalmente un fedelissimo dello "zar" Vladimir Putin in virtù dei comuni trascorsi nel KGB, "il secondo uomo più potente di Russia"; vedi J. Farchy, *Igor Sechin: Russia's second most powerful man*, http://www.ft.com/intl/cms/s/0/a8f24922-cef4-11e3-9165-00144feabdc0.html#axzz3GiBautfn.

seguito della crisi ucraina, è stata ventilata l'ipotesi di distacco dell'Internet russa dal resto della rete globale così come si starebbe valutando di imporre alle aziende straniere operanti in Russia l'obbligo di conservare i dati / documenti all'interno di *server* dislocati sul territorio della Federazione. Evidentemente, per il Cremlino, l'obiettivo di rafforzare la sovranità nazionale sul cyberspazio è questione di importanza vitale e si ottiene da un lato ponendo sotto la sua giurisdizione gli "archivi" che contengono le esistenze digitali dei propri cittadini e dall'altro controllando la rete fisica attraverso la quale esse potrebbero prendere la strada di Stati esteri "nemici". In particolare quest'ultima pare essere stata strutturata tenendo ben presenti tali esigenze: Andrei Soldatov, esperto di *intelligence* russa, interpellato dal Guardian osserva come l'operazione di "sconnessione", per quanto al momento improbabile, sia fattibile dal punto di vista tecnico a causa del numero «sorprendentemente basso» di Internet Exchange Point[16].

Con i propositi delle autorità moscovite siamo già saliti al terzo ed ultimo livello, vale a dire quello relativo ai rapporti tra Stati ed al ruolo strategico svolto dai *data center* / archivi in questo ambito: il processo di concentrazione "documentaria" al loro interno fa infatti sì che chi controlla i DC controlla, *de facto*, le informazioni, vale a dire il bene più prezioso (in termini economici, militari, sociali, etc.) della società contemporanea.

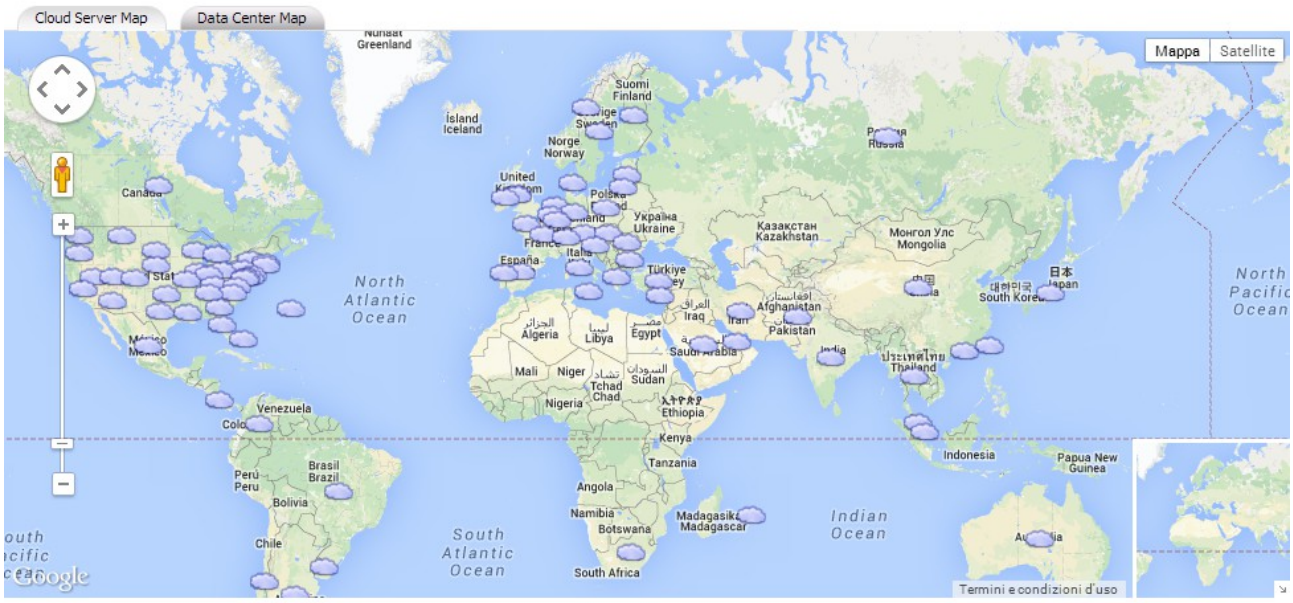
Come ripetutamente evidenziato, l'atteggiamento dei vari governi nei confronti di tali infrastrutture è, nel complesso, caratterizzato da una scarsa coerenza: da una parte se ne riconosce l'importanza, dall'altra si rinuncia a costruirne.

Al netto di questa "ambiguità" nei comportamenti, alcune valutazione complessive sono comunque possibili. In primo luogo, come si evince dalla mappa dei *cloud server* sparsi per il pianeta, la maggior parte di essi sono collocati all'interno degli Stati Uniti[17]: se tale situazione indubbiamente riflette in modo "naturale" il primato mondiale di Washington nei vari campi, dall'altro appare evidente che da tale situazione di predominio sulle informazioni essi cerchino di trarre vantaggio, anche (ma non solo) attraverso attività di spionaggio.

---

[16] Sull'argomento si vedano L. Harding, *Putin considers plan to unplug Russia from the Internet "in a emergency"*, http://www.theguardian.com/world/2014/sep/19/vladimir-putin-plan-unplug-russia-internet-emergency-kremlin-moscow e P. Sonne - O. Razumovskaya, *Russia Steps Up New Law to Control Foreign Internet Companies*, http://online.wsj.com/articles/russia-steps-up-new-law-to-control-foreign-internet-companies-1411574920.

[17] Per una navigazione interattiva si veda: http://www.datacentermap.com/cloud.html; si tenga presente che i dati riportati sono conferiti su base volontaria e dunque indicativi di una tendenza piuttosto che rappresentativi di valori esatti.



La constatazione di questa impostazione americanocentrica ci porta ad una ulteriore riflessione: noi tutti siamo portati a pensare, per definizione, alla Rete come ad una costruzione policentrica e distribuita. In verità ciò può valere al massimo per la parte fisica dei cavi (ed anche questo fino ad un certo punto, come si è dimostrato in altra sede; cfr. n. 11) giacché i gangli, vale a dire i *data center* che fungono da punti di accumulo e senza dei quali non ci sarebbero *bit* da far circolare, sono inesorabilmente dislocati negli Stati più progrediti (l'Africa ne è praticamente priva). L'esito finale è la riproposizione del consunto modello centro-periferia e la constatazione che la Rete, al di là delle apparenze, è meno "democratica" di quel che vorrebbe far credere.

Sempre dall'analisi della mappa balza agli occhi come in seconda posizione nella particolare classifica di chi ospita il maggior numero di archivi / *data center* si piazzi l'Unione Europea; quest'ultima, in virtù della particolare attenzione da sempre dimostrata nei confronti di tematiche quali la tutela della *privacy*, potrebbe farsi promotrice di una politica *ad hoc* (che a ben vedere è a tutti gli effetti politica archivistica) con l'obiettivo di realizzare un modello alternativo a quello statunitense e capace di smussarne gli aspetti più certi estremi.

In particolare, posto che a monte sarebbe necessaria un'armonizzazione delle varie legislazioni nazionali in materia di dati personali (tema sul quale Bruxelles è al lavoro), si potrebbe ipotizzare a livello comunitario la realizzazione di *data center* di dimensioni analoghe agli omologhi nordamericani ma rispettosi degli standard europei in tema di diritti. A tali archivi, destinati ad essere posti al servizio della *cloud*, ne andrebbero affiancati di minori dislocati a livello di regione / area metropolitana e deputati, questa volta, all'erogazione di quei servizi di tipo *fog* (IoT, *wearable*

computing) ritenuti essenziali alla realizzazione di quelle *smart city* che riprendono e rinnovano la tradizione urbana del Vecchio Continente[18].

I vantaggi derivanti da una siffatta ripartizione sarebbero molteplici: in primo luogo si andrebbe a riequilibrare, distribuendo a seconda dei casi dati e documenti a livello sovranazionale o locale, lo sbilanciamento attuale che li vede risiedere in maniera massiccia all'interno di *server* fuori non solo dal nostro controllo ma pure da quello dei nostri Stati di residenza.

In secondo luogo la ridondanza delle *location* fisiche, unita alla maggior distribuzione geografica, darebbe superiori garanzie in termini di sopravvivenza (*business continuity* / *disaster recovery*) e permetterebbe nel contempo di razionalizzare gli interventi, ottenendo sufficienti economie di scala.

In terzo luogo lo Stato (sovranazionale nello specifico) da un lato si riapproprierebbe della vitale funzione archiviale, con tutto ciò che ne consegue in termini di autolegittimazione, ma dall'altra la condividerebbe con altri soggetti "locali" che, impedendo l'accumularsi di "troppi dati" in un ristretto numero di DC, verrebbero a svolgere una delicatissima funzione di contrappeso: così facendo, questo perlomeno è l'auspicio, il sistema potrebbe assumere una dimensione più "umana", in modo da disinnescare potenziali conflitti per il dominio dell'informazione.

La questione pertanto, e qui concludiamo, a dispetto delle premesse tecnologiche viene ad assumere una dimensione "etica" che chiama in causa in modo pressoché "naturale" l'intera comunità archivistica, tale è stato l'impegno da essa da sempre profuso per queste tematiche. Il compito, evidentemente, non è facile dati gli attori in campo ma la speranza che, diversamente dal passato, i suoi consigli ed i suoi accorati appelli vengano raccolti dai nostri decisori politici, rimane.

[18] Per alcune riflessioni sul nesso esistente tra archivi e *smart city* si rimanda al proprio *Dalla città ideale alla smart city. Una riflessione sul ruolo degli archivi*, http://www.ilmondodegliarchivi.org/index.php/studi/item/303-dalla-citt%C3%A0-ideale-alla-smart-city-una-riflessione-sul-ruolo-degli-archivi.