

Verso una nuova archiveconomia: alcune riflessioni sull'evoluzione della disciplina nella transizione dall'analogico al digitale

Stefano Allegrezza^(a)

a) University of Udine, Italy, <http://orcid.org/0000-0002-7319-2483>

Contact: Stefano Allegrezza, stefano.allegrezza@uniud.it.

Received: 05 October 2016; **Accepted:** 11 November 2016; **First Published:** 15 January 2017

ABSTRACT

The paper aims at proposing some considerations on the evolution of the discipline of archival economy in the transition from 'traditional' to digital archives, focusing in particular on issues related to the use of new prefixes for binary multiples required for correctly estimate the size of digital archives and the capacity of storage systems. In particular, the paper aims at clarifying this particular aspect, which is rarely taken in consideration while it is very important to take it in consideration during the creation of a digital archive. As a consequence, the skills required to archivists who want to deal with digital archives must change and there is the need of a 'new' archival economy which combines the traditional topics with the issues related to the management of digital archives.

KEYWORDS

Archival economy; Digital archives; Datacentre; Storage; Binary prefix.

CITATION

S. Allegrezza. "Verso una nuova archiveconomia: alcune riflessioni sull'evoluzione della disciplina nella transizione dall'analogico al digitale". *JLIS.it* 8, 1 (January 2017): 114-126. doi: [10.4403/jlis.it-12140](https://doi.org/10.4403/jlis.it-12140).

Introduzione

Tradizionalmente l'archivonomia è definita come quella branca dell'archivistica che si occupa della disposizione e della custodia del materiale documentario, fornendo indicazioni sulla scelta dell'edificio e dei locali, sulle strutture entro le quali collocare il materiale (scaffalature, armadi, faldoni, cartelle, ecc.), sulle condizioni ambientali più adatte per la loro conservazione (temperatura, umidità, illuminazione, polvere, ecc.), sulle misure da adottare nei confronti delle calamità naturali (incendi, alluvioni, terremoti, ecc.) e contro il biodeterioramento (funghi, muffe, animali nocivi, insetti, parassiti, ecc.), fino ad arrivare ai piani di emergenza e ai piani di recupero dopo l'emergenza¹.

Il vocabolario Treccani definisce l'archivonomia come:

la disciplina che costituisce un settore dell'archivistica e tratta della tenuta e dell'amministrazione degli archivi; in particolare comprende le regole e gli accorgimenti per la costruzione dei fabbricati destinati ad accogliere gli archivî, la sistemazione degli scaffali e la preservazione del materiale archivistico dai pericoli di incendi, tarli, muffa, ecc.²

Eugenio Casanova (1928, 25-26), nel suo celebre manuale scrive:

noi possiamo distinguere nell'archivistica tre grandi divisioni della materia in discussione, che, secondo noi, sono le seguenti:

1. l'amministrazione generale esterna dell'archivio, e come dicono i tedeschi, con una sola parola, l'archivonomia;
2. l'ordinamento interno degli atti, ossia archivistica pura;
3. il servizio e natura giuridica dell'archivio.

Adolf Brenneke, nell'introduzione al suo trattato di Archivistica (1968, 21) fa ancora riferimento all'opera del suo illustre predecessore e scrive che:

il Casanova [...] suddivide l'archivistica in: archivistica pratica o tecnologica (archivonomia), la quale ha come fine principale quello di elaborare le basi scientifiche del trattamento esterno del materiale archivistico (ricezione, custodia, conservazione e restauro); archivistica propriamente detta, ossia archivistica teorica (archivistica pura), che studia le regole per la sua utilizzazione e quindi i metodi per l'ordinamento interno e si occupa della natura e delle varie specie di archivi; diritto archivistico, al quale si aggiunge, in fine, anche la storia degli archivi e dell'archivistica (Brenneke 1968, 21).

Secondo Virgilio Giordano, l'archivonomia è:

¹ Per una trattazione approfondita di questi argomenti si rimanda al volume di Bertini (2009). Si segnala anche il volume Bertini e Petrilli (2014).

² Cfr. la voce "archivonomia" sulla versione on line del vocabolario Treccani, <http://www.treccani.it/vocabolario/archivonomia>.

il ramo dell'archivistica che studia e suggerisce i migliori accorgimenti sia per le costruzioni e la strutturazione edilizia dei locali di archivio, sia per il loro arredamento funzionale e per la salvaguardia dei documenti dagli agenti distruttivi e dagli incendi (Giordano 1978, 163).

Lo stesso autore aggiunge che l'archiveconomia “può distinguersi in tre branche: l'edilizia, l'arredamento e la difesa dagli agenti distruttivi” (ibid.).

Più recentemente, Beatrice Ramazio, distinguendo tra archivistica teorica, archiveconomia, legislazione archivistica, archivistica informatica e archivistica speciale, definisce l'archiveconomia come:

quel settore dell'archivistica che si occupa dell'ubicazione, della struttura e degli elementi indispensabili nella costruzione dei locali da adibire ad archivio, oltre alla sistemazione delle scaffalature, al condizionamento dei documenti e a tutti gli accorgimenti da adottare per una collocazione e movimentazione delle risorse finalizzata alla loro conservazione permanente (Ramazio 2014, 97).

L'autrice specifica come “tale disciplina sia a tutti gli effetti un sapere trasversale, che coinvolge numerose competenze e professionalità” (Ramazio 2014, 102) riconoscendo l'importanza di un approccio multidisciplinare che deve giovare dell'apporto di più figure professionali.

Nel passaggio dagli archivi ‘tradizionali’ agli archivi digitali alcuni di questi concetti rimangono invariati, altri devono essere integralmente ‘rivisti’ in virtù del fatto che si ha a che fare non più con documenti cartacei ma con documenti digitali, i quali hanno una ben diversa materialità fatta non più di cellulosa o pelle di agnello ma di sequenze più o meno lunghe di bit ‘0’ ed ‘1’ registrati su uno dei tanti supporti di memorizzazione oggi esistenti.

Ma andiamo con ordine. Come è noto, i documenti che riguardano uno stesso affare formano un fascicolo e, nel mondo analogico, vengono collocati all'interno di una camicia o coperta in carta forte³ secondo l'ordine di archiviazione; sulla camicia “vanno riportati i dati identificativi del fascicolo: l'anno di apertura, il titolo e la classe nel cui ambito si apre il fascicolo, il numero progressivo attribuito al fascicolo [...], l'oggetto del fascicolo” (Bonfiglio-Dosio 2010, 68). Più fascicoli sono generalmente riuniti in un contenitore che nelle diverse zone d'Italia “possono chiamarsi con nomi diversi: cartelle, buste, mazzi, faldoni” (Bertini 2009, 110). Il fascicolo costituisce l'unità archivistica, ovvero l'unità di base indivisibile di un archivio. La busta è l'unità di condizionamento ed è spesso utilizzata, insieme al metro lineare, per indicare la consistenza quantitativa dell'archivio. Le buste vengono tipicamente collocate sui ripiani di scaffali, armadi o strutture.

Passando al mondo digitale, i documenti cartacei vengono ‘rimpiazzati’ da documenti informatici, che, in ultima analisi, sono *file* – e quindi sequenze di bit – e come tali necessitano, anche solamente per ‘esistere’, di un supporto su cui essere memorizzati. I fascicoli diventano ‘cartelle’ o altre forme di aggregazione documentale nel *file system* in uso, mentre il concetto di unità di condizionamento,

³ Si veda il Regio Decreto 25 gennaio 1900, n. 35 “Approvazione del regolamento per gli Uffici di registratura e di archivio delle Amministrazioni centrali”. L'art. 34 così recita: “Chiamasi fascicolo la riunione ordinata per data o per numero degli atti ricevuti e spediti pel medesimo affare” mentre il successivo art. 35 specifica le caratteristiche della coperta: “Ogni fascicolo ha una coperta di carta forte, di colore diverso per le diverse classi, alta centimetri 36, larga 28 [...]”.

stante la diversa ‘fisicità’ dei materiali digitali, perde molta della sua importanza. Gli scaffali scompaiono per lasciare il posto ai *rack*, strutture metalliche su cui vengono collocati i dispositivi di *storage* – come gli *hard disk* o i solid state disk (SSD) che stanno progressivamente soppiantando i primi – e l’infrastruttura necessaria per il loro funzionamento e che, nel loro complesso, consentono di memorizzare i documenti informatici. I locali degli archivi tradizionali vengono rimpiazzati dai *datacenter* che hanno anch’essi la necessità di adottare tutta una serie di misure di sicurezza sia logiche che fisiche per garantire la conservazione degli archivi digitali in essi ospitati al pari dei loro equivalenti analogici.

Come si può desumere da queste prime e rapide osservazioni, nella transizione dall’analogico al digitale vi sono molte differenze ma anche alcune somiglianze. Certamente vi sono operazioni come quelle di movimentazione del materiale documentario che hanno una minore fisicità, dal momento che sono normalmente compiute utilizzando strumenti informatici e senza alcuna attività manuale; tuttavia non si deve commettere l’errore di pensare che gli archivi digitali non abbiano una propria materialità e che quindi non necessitino di locali ed attrezzature idonei per essere ospitati: al contrario, esattamente come gli archivi ‘tradizionali’, anche gli archivi digitali hanno bisogno di locali progettati e realizzati nel rispetto dei più elevati standard di sicurezza.⁴ La fragilità dei bit è tale che bisogna adottare tutti i possibili accorgimenti se si vuole assicurare la preservazione nel tempo degli archivi digitali.

Non è questa la sede per affrontare in maniera esaustiva tutte le numerose questioni che andrebbero analizzate per comprendere appieno le dimensioni di questo cambiamento di paradigma (si pensi alla salvaguardia dei nuovi ‘supporti scrittori’ dai fattori di degrado legati alla loro concreta fisicità; alla sicurezza fisica e logica dei depositi digitali; alle esigenze di ridondanza; etc.) e che, per evidenti motivi di brevità, non possono trovare spazio in questa sede; tuttavia, è utile approfondire alcune questioni che, come si vedrà, possono avere delle conseguenze importanti per la custodia degli archivi.

Uno degli aspetti che occorre prendere in considerazione durante la fase di formazione di un archivio – sia esso tradizionale o digitale – e che rientra a pieno titolo nel campo di applicazione dell’archivologia, è quello relativo alla determinazione della quantità di spazio necessaria per accogliere i fondi. Nel caso degli archivi digitali ciò si traduce nella stima della dimensione ‘informatica’ dell’archivio che si andrà a costituire e, di conseguenza, della capacità di memorizzazione (o di *storage*) necessaria. Una stima corretta è molto importante perché presenta risvolti anche dal punto di vista economico, oltre che organizzativo: se la stima è per difetto, è possibile che il sistema di *storage* non sia in grado di accogliere tutti i documenti e le altre informazioni (come i metadati ad essi associati) che si ha la necessità di conservare e quindi si debba ricorrere all’acquisizione in un secondo momento, di ulteriore capacità di *storage*; se la stima è per eccesso, è possibile che si vada incontro a costi che potevano essere evitati con una stima più oculata.⁵ Nel seguito tratteremo in

⁴ Si pensi, ad esempio, agli standard per la sicurezza della serie ISO 27000, oppure a quelli sulla gestione della continuità operativa (ISO 22301:2012) od ancora alla verifica dell’affidabilità dei depositi digitali secondo lo standard ISO 16363. Per un approfondimento su queste tematiche si vedano, tra gli altri: Guercio (2013); Pigiapoco (2010).

⁵ Come si vedrà più avanti, quando le capacità di memorizzazione sono dell’ordine del *petabyte* o dell’*exabyte*, commettere errori nello stimare le capacità necessarie può avere ripercussioni dal punto di vista economico di non poco conto.

profondità questo specifico argomento che risulta importante e spesso non pienamente valutato tutte le volte che si progetta la costituzione di un archivio digitale.

La misura della ‘consistenza’ degli archivi digitali

Quando si gestiscono archivi digitali l’unità di misura utilizzata per misurarne la consistenza non coincide, com’è ovvio, con quelle tradizionalmente utilizzate a questo scopo nel mondo analogico (il metro lineare, il numero di buste, ecc.). Normalmente si parla di ‘dimensione’ dell’archivio (o di *file* e cartelle) e di ‘capacità di memorizzazione’ dei supporti di archiviazione e l’unità di misura utilizzata è il *byte*.⁶ Tuttavia, date le grandezze in gioco, ormai da decenni tali misure non vengono più espresse in termini di semplici *byte* (come avveniva trenta anni or sono, quando le dimensioni in gioco erano molto inferiori)⁷ ma utilizzando i suoi multipli. Oggi termini come *kilobyte*, *megabyte*, *gigabyte*, *terabyte*, ecc. sono entrati nel lessico comune⁸ ed è ormai noto che i prefissi ‘kilo’, ‘mega’, ‘giga’, ‘tera’ ed i corrispondenti simboli ‘k’, ‘M’, ‘G’, ‘T’ hanno il valore, rispettivamente, di 2^{10} (= 1024), 2^{20} (= 1024^2), 2^{30} (= 1024^3), 2^{40} (= 1024^4) (ogni prefisso vale esattamente 1024 volte il precedente).

Purtroppo, questi simboli (ed i successivi ‘P’, ‘E’, ‘Z’, ‘Y’ di cui si tratterà più avanti) sono uguali a quelli previsti dal Sistema Internazionale di unità misura (S.I.), il quale, però, attribuisce loro significati differenti. In particolare, nel S.I., il simbolo ‘k’ corrisponde a 10^3 (= 1000), il simbolo ‘M’ a 10^6 (= 1000^2), il simbolo ‘G’ a 10^9 (1000^3) e il simbolo ‘T’ a 10^{12} (= 1000^4); ogni multiplo vale quindi esattamente 1000 volte il precedente.⁹ La differenza, come si può facilmente intuire, non è trascurabile. Ad esempio, 1 kB nell’informatica non è pari a 1.000 *byte* ma a 1.024 *byte*, con una differenza del 2,4%; allo stesso modo 1 MB non è pari a 1.000.000 di *byte* ma a 1.048.576 *byte*, con una differenza, questa volta, del 4,9%. Lo scostamento tra le capacità di memorizzazione espresse nei due sistemi di numerazione (binario e decimale) cresce all’aumentare delle grandezze in gioco fino a raggiungere valori assolutamente non trascurabili (si veda la Tabella 1 più avanti).

Tutto questo porta ad una grande confusione riguardo il significato dei prefissi utilizzati. Ad esempio, in assenza di altre informazioni, non è chiaro quale sia la capacità di memorizzazione di un *hard disk* da 500 GB. Si tratta di 500×1024^3 *byte* o di 500×1000^3 *byte*? Quale delle due è la capacità di memorizzazione corretta? Passando alle dimensioni dei *file*, ci potremmo chiedere quale sia la dimensione di un file da 2 GB: si tratta di 2×1024^3 *byte* o 2×1000^3 *byte*? Anticipando risposte che risulteranno evidenti più avanti, possiamo dire che il prefisso ‘G’ assume nel primo caso il valore di 1000^3 mentre nel secondo caso il valore è pari a 1024^3 . Non c’è dubbio che il fatto che lo stesso prefisso

⁶ Com’è noto, un *byte* corrisponde ad una sequenza di otto bit. Per convenzione il *byte* è indicato con la lettera ‘B’ (maiuscola), mentre il bit è indicato con la lettera ‘b’ (minuscola).

⁷ Prima del rilascio dei primi sistemi Macintosh avvenuto nel 1984, le dimensioni dei *file* erano solitamente indicate dal sistema operativo senza l’impiego di alcun prefisso. Ad esempio, i sistemi operativi dell’epoca (come il ben noto MS-DOS) avrebbero indicato la dimensione di un file di 30 kB come ‘30.720 *byte*’ (30 kB = 30 x 1024 *byte*) senza far uso di alcun prefisso.

⁸ Nel gergo comune si usano quasi sempre i nomi abbreviati di ‘mega’, ‘giga’ e ‘tera’, creando non poca confusione. Ad esempio, le inserzioni pubblicitarie reclamizzano *pendrive* USB da ‘32 giga’, linee ADSL da ‘20 mega’, e così via, senza specificare l’unità di misura (nel primo caso si tratta di 32 GB (gigabyte), nel secondo di 20 Mb (megabit)).

⁹ La differenza è dovuta al fatto che il S.I. definisce i multipli (ed i sottomultipli) basandosi sulle potenze del dieci (come 10^3 , 10^6 , ecc.), mentre nell’informatica si utilizzano multipli che sono basati sulle potenze del due (come 2^{10} , 2^{20} , ecc.).

(‘G’ in questo caso) assuma significati diversi pur trattandosi dello stesso contesto (ovvero quello informatico) genera una grande confusione.

Ma come si è potuto arrivare a ciò? L’usanza di utilizzare i prefissi per i multipli decimali (quelli stabiliti dal Sistema Internazionale) per indicare le capacità di memorizzazione risale agli inizi degli anni ‘60. A quel tempo, le capacità in gioco superavano a malapena le poche decine di *kilobyte*, e gli informatici notarono che la decima potenza del 2 (il numero $2^{10} = 1024$) era quasi uguale alla terza potenza del 10 (il numero $10^3 = 1000$) e, puramente per una questione di convenienza, iniziarono a far riferimento a 1.024 *byte* come ad ‘1 kB’, utilizzando il prefisso ‘k’ del S.I.¹⁰. Dopotutto la differenza era solo del 2,4% ed inoltre tutti i professionisti informatici sapevano che “nel loro ambiente” il prefisso ‘k’ significava 1.024 e non 1.000. Nonostante la sua imprecisione, quest’uso improprio del prefisso ‘k’ si estese al di fuori dell’ambito informatico e diventò di uso comune. Con il passare del tempo, le capacità di memorizzazione delle memorie di massa dei computer cominciarono ad essere sempre più grandi: dai *kilobyte* si passò ai *megabyte*, poi ai *gigabyte*, quindi ai *terabyte* e così via, in una corsa inarrestabile che sembra non avere più fine. Oggi non è raro trovare sistemi di grandi dimensioni nei quali le capacità di memorizzazione sono dell’ordine del *petabyte* o dell’*exabyte*.¹¹ Ovviamente quando si vanno a considerare capacità di memorizzazione di quest’ordine di grandezza, la differenza tra il valore espresso con i multipli decimali e l’equivalente binario non è più trascurabile. Ad esempio, se si considerano grandezze dell’ordine del *tera*, la differenza tra il valore decimale (10^{12}) e l’equivalente binario (2^{40}) si avvicina al 10% (si veda la Tabella 3) e non è certamente trascurabile. Se si prosegue oltre, ci si accorge che la differenza aumenta ancor di più: ad esempio se l’ordine di grandezza è quello del *peta* la differenza è intorno al 12%; passando all’*exa* la differenza si avvicina al 15%, e così via; più aumenta l’ordine di grandezza, più aumenta la differenza. Questa è la ragione principale che ha portato diversi organismi di normazione a definire i nuovi prefissi per i multipli binari.

I prefissi per i multipli decimali

Prima di affrontare l’argomento dei ‘nuovi’ prefissi per i multipli binari, è opportuno ricordare quali sono i prefissi utilizzati per i multipli decimali. I prefissi per i multipli decimali sono stati definiti dal Sistema Internazionale di unità di misura (S.I.). Il S.I. è il più utilizzato tra i sistemi di unità di misura¹² e, assieme al Sistema C.G.S., viene spesso indicato come sistema metrico, soprattutto nei paesi

¹⁰ Si tenga presente che nella rappresentazione dell’aritmetica binaria (che fa uso delle due sole cifre 0 e 1), il numero 1000 corrisponde a ‘1111101000’, mentre il numero 1.024 corrisponde a ‘1000000000’. È quindi quest’ultima, in realtà, la ‘cifra tonda’, mentre l’uso del numero 1.000 come multiplo dell’unità di memoria avrebbe comportato molte complicazioni e pochissimi vantaggi.

¹¹ Basta pensare ai datacenter di colossi quali Google, Amazon, Microsoft, Facebook che gestiscono archivi digitali con capacità di memorizzazione dell’ordine dell’*exabyte* e stanno velocemente avvicinandosi a grandezze dell’ordine dello *zettabyte*. Si vedano, a tal proposito, gli articoli “Facts and Stats” 2013 e Miller 2013.

¹² Il Sistema Internazionale (S.I.) di unità di misura è stato definito anche nella norma ISO 1000:1992 “*SI units and recommendations for the use of their multiples and of certain other units*” (attualmente la norma è stata rivista ed è confluita nella nuova ISO 80000-1:2009 “*Quantities and units -- Part 1: General*”). Questo sistema di grandezze ed unità di misura nacque nel 1889 con la prima Conferenza Generale dei Pesi e delle Misure e prese il nome di ‘Sistema MKS’ perché comprendeva solo le unità fondamentali di lunghezza (*metro*), massa (*chilogrammo*) e tempo (*secondo*). Oggi il S.I. è basato su sette grandezze fondamentali (metro, chilogrammo, secondo, ampere, kelvin, candela, mole) a partire dalle quali vengono definite tutte le grandezze derivate.

anglosassoni. Le unità, la terminologia e le raccomandazioni del Sistema internazionale vengono stabilite dalla Conferenza Generale dei Pesi e delle Misure (CGPM), organismo collegato con l'Ufficio internazionale dei pesi e delle misure (*International Bureau of Weights and Measures*, BIPM). Il S.I. definisce una serie di prefissi da premettere alle unità di misura per identificare i loro multipli (si veda la Tabella 2).

Può essere utile conoscere l'etimologia di questi prefissi:

- 'kilo' (k)¹³ deriva dal greco *khilia* e significa, abbastanza curiosamente, 'mille'; si tratta dell'unico prefisso con un significato numerico diretto.

I successivi tre prefissi hanno un significato descrittivo o mitologico:

- 'mega' (M) deriva dal greco *megas* e ha il significato di 'grande';
- 'giga' (G) deriva dal Latino *gigas*, che deriva a sua volta dal greco ed ha il significato di 'gigante';
- 'tera' (T) viene dal greco *teras* e significa 'mostro', prodigio, miracolo, cosa terribile o straordinaria.

Prefisso	Simbolo	Valore	Significato
yotta	Y	$10^{24} = 1000^8$	quadrilione
zetta	Z	$10^{21} = 1000^7$	triliardo
exa	E	$10^{18} = 1000^6$	trilione
peta	P	$10^{15} = 1000^5$	biliardo
tera	T	$10^{12} = 1000^4$	bilione
giga	G	$10^9 = 1000^3$	miliardo
mega	M	$10^6 = 1000^2$	milione
chilo	k	$10^3 = 1000^1$	mille

Tabella 3. I prefissi per i multipli decimali.

I successivi due prefissi fanno riferimento ai numeri:¹⁴

- 'peta' (P) deriva dal greco *pente*, che significa 'cinque' (con riferimento a 1000^5);

¹³ Si noti che va utilizzato il prefisso 'k' (nella forma minuscola anziché maiuscola) dal momento che il simbolo 'K' viene utilizzato nel S.I. come simbolo del grado Kelvin, unità di misura della temperatura. Ad esempio, è errato indicare il chilogrammo come 'Kg' mentre si dovrebbe scrivere 'kg'. Allo stesso modo il chilometro non va indicato come 'km' e non come 'Km'.

¹⁴ I prefissi 'peta' ed 'exa' furono aggiunti nel 1975 dalla Conferenza Generale dei Pesi e delle Misure; i successivi prefissi 'zetta' e 'yotta' furono aggiunti nel 1990.

- ‘exa’ (E) deriva dal greco *exi* che significa ‘sei’ (con riferimento a 1000⁶).

Gli ultimi due prefissi fanno riferimento all’alfabeto latino:

- ‘zetta’ (Z) è l’ultima lettera dell’alfabeto latino e deriva dal latino *septem*, con il significato di ‘sette’ (con riferimento a 1000⁷);
- ‘yotta’ (Y) è la penultima lettera dell’alfabeto latino e deriva dal greco *okto*, con il significato di ‘otto’ (con riferimento a 1000⁸).

I nuovi prefissi per i multipli binari

Come già anticipato, l’abuso da parte degli informatici dei prefissi decimali non per indicare i multipli decimali (per cui sono stati pensati e standardizzati) ma i multipli binari, sfruttando il fortuito caso che la decima potenza di due ($2^{10} = 1.024$) è molto vicina alla terza potenza del dieci ($10^3 = 1.000$), ha spinto vari organismi di normazione a stabilire dei nuovi prefissi, insieme con i relativi simboli, da utilizzare per i multipli binari al posto dei prefissi decimali. In particolare, nel 1999 l’International Electrotechnical Commission (IEC) ha pubblicato un’appendice alla norma IEC 60027-2,¹⁵ che stabilisce l’utilizzo dei nuovi prefissi ‘kibi’, ‘mebi’, ‘gibi’, ‘tebi’, ‘pebi’, ‘exbi’ e dei corrispondenti simboli ‘Ki’,¹⁶ ‘Mi’, ‘Gi’, ‘Ti’, ‘Pi’, ‘Ei’ per specificare i multipli binari ed eliminare ogni ambiguità. I nomi dei prefissi sono ricavati dalle prime due lettere dei corrispondenti prefissi del S.I. seguiti dalla desinenza ‘bi’ (abbreviazione di ‘binary’). Ad esempio *kibi* significa ‘kilobinary’ e corrisponde a $2^{10} = 1024$, *mebi* significa ‘megabinary’ e corrisponde a $2^{20} = 1024^2$, e così via. I prefissi raccomandati dalla norma raccomandano riepilogati nella Tabella 2.¹⁷

Prefisso	Simbolo	Valore	Significato
yobi	Yi	$2^{80} = 1024^8$	yottabinary
zebi	Zi	$2^{70} = 1024^7$	zettabinary
exbi	Ei	$2^{60} = 1024^6$	exabinary
pebi	Pi	$2^{50} = 1024^5$	petabinary

¹⁵ La norma IEC 60027 “*Letter symbols to be used in electrical technology*”, precedentemente denominata IEC 27, è composta da sei parti; la parte che descrive i simboli da utilizzare come multipli binari è la seconda (IEC 60027-2), adesso sostituita dalla norma IEC 80000-13:2008: “*Quantities and units-- Part 13: Information science and technology*”. La norma IEC 60027-2 stabilisce che, dal momento della sua pubblicazione, i prefissi del S.I. dovranno avere esclusivamente il significato decimale e non più anche binario. Quest’appendice aveva definito i prefissi solo fino a ‘exbi’, ma la terza edizione della norma, pubblicata nel 2005, ha aggiunto i prefissi ‘zebi’ e ‘yobi’, insieme con i relativi simboli ‘Zi’ e ‘Yi’, completando in questo modo l’analogia tra i prefissi decimali standard ed i loro omologhi binari

¹⁶ Si noti che, per coerenza con gli altri prefissi, l’IEC ha stabilito l’utilizzo della lettera ‘K’ maiuscola nel prefisso ‘Ki’, mentre per l’analogo prefisso il S.I. utilizza la lettera ‘k’ minuscola. È pertanto errato scrivere ‘KB’ per indicare 1.000 *byte* (si deve scrivere ‘kB’) mentre è corretto scrivere KiB per indicare 1.024 *byte*.

¹⁷ I nuovi prefissi possono essere utilizzati sia con il termine ‘byte’ che con ‘bit’. Ad esempio per indicare 2^{10} *byte* = 1.024 *byte* si usa il termine *kibibyte*, abbreviazione di ‘kilobinary byte’. Analogamente con *mebibyte* si indica 2^{20} *byte* = 1024² *byte*. Invece, con il termine *kibibit*, abbreviazione di ‘kilobinary bit’, si indica 2^{10} *bit* = 1024 *bit*.

Prefisso	Simbolo	Valore	Significato
tebi	Ti	$2^{40} = 1024^4$	terabinary
gibi	Gi	$2^{30} = 1024^3$	gigabinary
mebi	Mi	$2^{20} = 1024^2$	megabinary
kibi	Ki	$2^{10} = 1024^1$	kilobinary

Tabella 4. I prefissi per i multipli binari.

Vediamo alcuni esempi di equivalenze:

- un KiB è uguale a 1024 byte, mentre un kB è uguale a 1000 byte;
- un MiB è uguale a 1024^2 byte, mentre un MB è uguale a 1000^2 byte;
- un GiB è uguale a 1024^3 byte, mentre un GB è uguale a 1000^3 byte;
- un TiB è uguale a 1024^4 byte, mentre un TB è uguale a 1000^4 byte.

I nuovi prefissi binari sono stati successivamente riconosciuti anche da altri organismi di standardizzazione, tra cui l'IEEE, che ha pubblicato al riguardo la norma IEEE 1541-2002 "*IEEE Standard for Prefixes for Binary Multiples*".¹⁸ Nel 1998 l'Ufficio internazionale dei pesi e delle misure ha pubblicato un documento in cui affermava, tra le altre cose, che i prefissi del S.I. debbono riferirsi solamente alle potenze del dieci e non possono essere utilizzati per indicare i multipli binari. I prefissi binari sono stati adottati anche dal CENELEC con la norma armonizzata HD 60027-2:2003-03 e di conseguenza sono obbligatori anche nell'Unione Europea. Le norme ISO 31 e IEC 60027 sono state oggetto di revisione da parte delle due organizzazioni ISO e IEC: la norma armonizzata che ne è scaturita è denominata ISO/IEC 80000:2008 "*Quantities and units*" e recepisce anch'essa i nuovi prefissi binari (precisamente nella parte IEC 80000-13: *Information science and technology*).

Anche l'*International Bureau of Weights and Measures* (BIPM), che mantiene il Sistema Internazionale di unità di misura (S.I.), vieta espressamente l'utilizzo dei prefissi del S.I. per indicare multipli binari, dato che tra le grandezze del S.I. non figurano le 'unità di informazione' (mentre figurano altre grandezze, come la lunghezza, la massa, il tempo, l'intensità di corrente, ecc.) e raccomanda in tal caso l'utilizzo dei nuovi prefissi binari.

L'applicazione puntuale e rigorosa dei nuovi prefissi binari avrebbe dovuto eliminare o almeno ridurre la confusione attualmente presente, ma le cose non stanno ancora così. Questi nuovi prefissi faticano ad entrare nell'uso corrente e incontrano notevoli difficoltà nell'essere accettati. Una delle motivazioni principali che viene addotta è il fatto che l'utilizzo dei prefissi del S.I. per indicare i multipli binari è ormai una consuetudine ben radicata e che i principali sistemi operativi e le principali

¹⁸ La norma IEEE 1541 ricalca in maniera abbastanza fedele l'appendice alla norma IEC 60027-2, con l'unica eccezione che quest'ultima utilizza la parola 'bit' (anziché la lettera 'b') come simbolo per il *bit*.

applicazioni ancora continuano ad utilizzarli.¹⁹ In questo modo l'ambiguità e la confusione continuano, purtroppo, a persistere.

L'errore nell'indicazione delle dimensioni e delle capacità di memorizzazione

Sebbene a prima vista la differenza tra i valori corrispondenti ai prefissi binari e quelli corrispondenti ai prefissi decimali possa sembrare irrilevante, andando ad indagare meglio la questione si scopre che non è assolutamente così. Infatti, se è vero che per capacità di memorizzazione dell'ordine del kB l'errore è trascurabile, aumentando gli ordini di grandezza aumentano le differenze percentuali tra i valori binari e quelli decimali. Ad esempio, se si considerano capacità dell'ordine del TB, si ha:

$$1 \text{ TB} / 1 \text{ TiB} = 1000^4 \text{ B} / 1024^4 \text{ B} = 0,909$$

ossia un errore all'incirca del 10%.

Questo significa, ad esempio, che se un sistema di *storage* (si pensi anche semplicemente ad un comune *hard disk*) ha una capacità dichiarata dal costruttore di 500 GB, l'effettiva capacità di memorizzazione, espressa con i corretti prefissi binari, è pari a $500 \text{ GB} \times 0,909 = 454,5 \text{ GiB}$, ovvero all'incirca il 10% in meno. Si tratta di una quantità che non è affatto trascurabile, e l'errore aumenta ancor di più man mano che si considerano sistemi di storage con capacità via via sempre più ampie.

Equivalenza	Errore	Equivalenza	Errore
1 YB = 0,827 YiB	17,3%	1 YiB = 1,209 YB	20,9%
1 ZB = 0,847 ZiB	15,3%	1 ZiB = 1,181 ZB	18,1%
1 EB = 0,867 EiB	13,3%	1 EiB = 1,153 EB	15,3%
1 PB = 0,888 PiB	11,2%	1 PiB = 1,126 PB	12,6%
1 TB = 0,909 TiB	9,1%	1 TiB = 1,100 TB	10,0%
1 GB = 0,931 GiB	6,9%	1 GiB = 1,074 GB	7,4%
1 MB = 0,954 MiB	4,6%	1 MiB = 1,049 MB	4,9%
1 kB = 0,976 KiB	2,4%	1 KiB = 1,024 kB	2,4%

Tabella 5. Prefissi decimali e binari ed errori.

La Tabella 6 mostra gli errori che si commettono utilizzando i prefissi decimali al posto di quelli binari e viceversa. Come si può vedere, se con grandezze dell'ordine del *kilobyte* la differenza è del 2,4%, con grandezze dell'ordine del *terabyte* si sale a valori intorno al 9-10% e con grandezze dell'ordine dell'*exabyte* si giunge a valori intorno al 13-15%. Aumentando ulteriormente le grandezze in gioco (ad esempio, dell'ordine dello *yottabyte*), si arrivano a differenze che giungono fino al 21%. Si tratta di valori assolutamente non trascurabili e che devono essere noti a coloro che si occupano di archivi

¹⁹ In realtà, buona parte delle applicazioni *free* o *open source* (ad esempio, alcune delle distribuzioni basate su Linux) ha ormai adottato i nuovi prefissi.

digitali, dal momento che influiscono sulle scelte operate ed hanno un grande impatto anche sul versante economico, considerati i costi dei sistemi di *storage*.

Le implicazioni per gli archivi digitali

Gli argomenti fin qui presi in esame dovrebbero essere tenuti in conto da tutti coloro che si occupano di archivi digitali (archivisti informatici, records manager, responsabili della conservazione digitale, responsabili del Servizio per la tenuta del protocollo informatico, la gestione dei flussi documentali e degli archivi, ecc.) in tutte le fasi di vita di un archivio digitale e soprattutto nella fase della formazione.²⁰ In particolare, sono due i momenti in cui queste conoscenze risultano particolarmente utili. Un primo momento è quello della stima della *dimensione* dell'archivio. Come si è visto, i *file system* dei sistemi operativi utilizzano quasi unanimemente i prefissi decimali per indicare le dimensioni di *file*, cartelle, volumi, ecc., attribuendo però loro un valore binario. Chi si occupa di archivi digitali deve essere cosciente di questa discrepanza e deve saper stimare correttamente le dimensioni utilizzando i prefissi binari, in maniera da ottenere risultati congrui. Un secondo momento è quello della stima della *capacità di memorizzazione* dei sistemi di *storage*. In questo caso, chi si occupa di archivi digitali deve sapere che, in linea di massima, i costruttori esprimono le capacità di memorizzazione impiegando i prefissi decimali ed attribuendo loro un significato decimale; ma deve sapere che esistono anche delle eccezioni e quindi deve saper distinguere i vari casi che possono presentarsi.

Il seguente esempio, benché elementare e con una valenza puramente didattica, può servire a chiarire meglio la questione. Poniamo il caso di un archivista che sia chiamato a gestire un archivio digitale con una dimensione, rilevata dal *file system* del sistema operativo in uso, di 950 GB.²¹ Al momento della scelta del supporto idoneo per la copia di *backup* di tale archivio, egli effettua un'analisi di mercato e rileva che sono disponibili *hard disk* della dimensione di 1 TB e 2 TB (più costosi); ignorando completamente le questioni trattate nel presente articolo, egli si orienterà verso l'acquisto di un *hard disk* da 1 TB, ritenendolo sufficiente per le sue necessità di archiviazione.²² Ma le cose non stanno così: al momento di riversare il suo archivio sul supporto, avrà l'amara sorpresa di scoprire che la capacità di memorizzazione dell'*hard disk* da 1 TB, ovvero 1.000 GB, non è sufficiente per contenere 950 GB di dati! Questa situazione paradossale deriva dal fatto che per esprimere la dimensione dell'archivio e la capacità di memorizzazione del supporto sono stati utilizzati gli stessi prefissi ma con due significati diversi. Dov'è l'errore? Certamente non da parte del costruttore dell'*hard disk*, il quale ha scelto di indicarne la capacità di memorizzazione utilizzando i prefissi decimali (ed è un suo diritto, purché la capacità sia indicata correttamente secondo il prefisso utilizzato). L'errore è imputabile, in questo caso, al sistema operativo utilizzato che ha indicato le

²⁰ Le stesse considerazioni si rivelano molto importanti, in generale, anche nel caso di progetti di digitalizzazione nei quali le capacità di memorizzazione in gioco sono spesso assai elevate e non raramente sono dell'ordine del *terabyte* o del *petabyte*.

²¹ Potrebbe trattarsi, ad esempio, di un archivio formato da 950.000 documenti in formato PDF/A ciascuno della dimensione media di 1 MB.

²² Si noti che negli archivi tradizionali le cose non vanno poi così diversamente: normalmente l'acquisto delle scaffalature viene effettuato in base alla stima della consistenza del materiale documentario che deve essere collocato su di esse; in quel caso la misura è effettuata utilizzando i metri lineari o il numero dei 'pezzi'; in questo caso si utilizzano i byte ed i suoi multipli.

dimensioni di *file* e cartelle utilizzando prefissi decimali ma con un significato che è proprio dei prefissi binari: in altre parole, il sistema operativo ha fornito per la dimensione dell'archivio una misura errata, pari a 950 GB, mentre avrebbe dovuto indicare la dimensione corretta pari a 950 GiB.

In generale, per capire se un sistema di *storage* è sufficiente per 'contenere' un archivio digitale avente una determinata dimensione occorre effettuare le opportune conversioni ed esprimere tutte le grandezze in gioco mediante prefissi omogenei (o tutti decimali o tutti binari). Nell'esempio appena visto, se si vogliono esprimere tutte le grandezze utilizzando i prefissi decimali, si deve operare in questo modo: la capacità di memorizzazione dell'*hard disk*, pari ad 1 TB, è espressa correttamente e quindi non necessita di alcuna conversione; la dimensione dell'archivio, pari a 950 GB, deve invece essere 'interpretata' come 950 GiB; effettuando la conversione per esprimerla mediante i prefissi decimali si ottiene: $950 \text{ GiB} = 950 \times 1,074 \text{ GB} = 1.020 \text{ GB}$ ²³. Quindi, si può concludere che la dimensione dell'archivio supera di 20 GB la capacità di memorizzazione dell'*hard disk* che, pertanto, non è sufficiente a 'contenere' l'intero archivio.²⁴

L'esempio appena visto, sebbene molto semplice, può essere utile a comprendere quanto sia importante, per chi si occupa di archivi digitali, possedere una solida conoscenza di questi (ed altri) concetti di archiveconomia 'digitale'. L'archivista 'digitale' deve conoscere i prefissi per i multipli sia decimali che binari; essere in grado di utilizzarli per indicare in maniera corretta le dimensioni degli archivi e le capacità di memorizzazione dei supporti, effettuando eventualmente le necessarie conversioni; infine, deve essere in grado di comprendere se le indicazioni delle capacità di memorizzazione – che si trovano apposte sulle confezioni d'acquisto dei supporti di memorizzazione o che sono specificate nelle schede tecniche dei sistemi di storage – e le indicazioni delle dimensioni degli archivi fornite dal computer siano corrette oppure no, sapendo individuare e 'correggere' eventuali errori riscontrati sia nell'uno che nell'altro caso.²⁵

Considerazioni finali

Così come l'archiveconomia 'classica' tradizionalmente prendeva in considerazione, tra gli altri, aspetti quali la stima della consistenza dell'archivio e delle strutture necessarie ad accoglierlo, così la 'nuova' archiveconomia deve prendere in considerazione, tra gli altri, aspetti quali la determinazione delle dimensioni informatiche' degli archivi e la stima delle capacità di memorizzazione necessarie per contenerlo, utilizzando correttamente le unità di misura del mondo digitale.²⁶ Non c'è dubbio che l'archivista che voglia affiancare alle competenze necessarie per la gestione degli archivi tradizionali quelle necessarie per la gestione degli archivi digitali – una realtà con cui è ormai necessario

²³ Per l'equivalenza si faccia riferimento alla Tabella 3.

²⁴ È possibile ragionare anche nel modo opposto, ovvero esprimendo tutte le grandezze mediante i prefissi binari. In questo caso occorre convertire la capacità di memorizzazione dell'*hard disk* come segue: $1 \text{ TB} = 1000 \text{ GB} = 1000 \times 0,909 \text{ GiB} = 909 \text{ GiB}$. La dimensione dell'archivio è pari a 950 GiB effettivi, per cui si può concludere anche in questo caso che la dimensione dell'archivio supera la capacità di memorizzazione dell'*hard disk* da 1 TB.

²⁵ Nel caso dell'esempio precedente deve essere in grado di capire che la dimensione effettiva dell'archivio non è ciò che legge sul display del computer (ovvero 950 GB) ma 950 GiB; inoltre, deve essere in grado di calcolare l'equivalente espresso in GB (ovvero 1.020 GiB) per poter valutare se il supporto scelto è sufficiente oppure no per la sua memorizzazione.

²⁶ Ad esempio, per stimare la capacità di memorizzazione necessaria per accogliere le immagini acquisite in un progetto di digitalizzazione o quella necessaria per l'implementazione dell'archivio digitale di un'azienda o di un ente pubblico.

confrontarsi – dovrà avere conoscenze di questo tipo ed essere in grado di padroneggiare questi concetti. Di conseguenza, la formazione degli archivisti dovrà inevitabilmente essere rivista sulla base di queste mutate esigenze e sarà necessaria una rivisitazione di molti dei concetti tradizionali fino ad arrivare ad una ‘nuova’ archiveconomia, che affianchi ai tradizionali argomenti di studio le tematiche connesse con la gestione degli archivi digitali.

Bibliografia

- Bertini, Maria Barbara. 2009. *La conservazione dei beni archivistici e librari*. Roma: Carocci editore.
- Bertini, Maria Barbara, e Vincenza Petrilli. 2014. *I custodi della memoria: l'edilizia archivistica italiana statale del XX secolo*. Rimini: Maggioli Editore.
- Bonfiglio-Dosio, Giorgetta. 2010. *Primi passi nel mondo degli archivi. Temi e testi per la formazione archivistica di primo livello* (4 ed.). Padova: Cleup.
- Brenneke, Adolf. 1968. *Archivistica, contributo alla teoria ed alla storia archivistica europea* (traduzione italiana di Renato Perrella). Milano: Giuffrè.
<http://www.icar.beniculturali.it/biblio/pdf/Brenneke/brenneke.pdf>
- Casanova, Eugenio. 1928. *Archivistica* (2 ed.). Siena: Stab. Arti Grafiche Lazzari.
<http://www.icar.beniculturali.it/biblio/pdf/EuCa/totalCasanova.pdf>
- “Facts and Stats of World’s largest data centers”. 2013.
<https://storageservers.wordpress.com/2013/07/17/facts-and-stats-of-worlds-largest-data-centers>.
- Giordano, Virgilio. 1978. *Archivistica e beni culturali*. Caltanissetta-Roma: Salvatore Sciascia Editore.
- Guercio, Maria. 2013. *Conservare il digitale. Principi, metodi e procedure per la conservazione a lungo termine di documenti digitali*. Roma-Bari: Laterza.
- . 2010. *Archivistica informatica. I documenti in ambiente digitale* (nuova edizione). Roma: Carocci.
- Miller, Rich. 2013. “Facebook Builds Exabyte Data Centers for Cold Storage”.
<http://www.datacenterknowledge.com/archives/2013/01/18/facebook-builds-new-data-centers-for-cold-storage>.
- Pigliapoco, Stefano (a cura di). 2010. *Conservare il digitale*. Macerata: EUM.
- Ramazio, Beatrice. 2014. “Archivistica e archiveconomia”. In *I custodi della memoria: l'edilizia archivistica italiana statale del XX secolo* (a cura di Maria Barbara Bertini, e Vincenza Petrilli). Rimini: Maggioli editore.