

# Conservazione delle memorie digitali rischi ed emergenze sei casi di studio

A cura di  
Alessandra Ruggiero \*

## *INDICE*

PREFAZIONE di Luciano Scala

INTRODUZIONE di Alessandra Ruggiero

### I CASI DI STUDIO:

Censimento USA del 1960. Census Bureau / National Archives and Records Administration

Combat Air Activities File (CACTA) / National Archives and Records Administration

Banca dati del Consorzio Neapolis / Soprintendenza Archeologica di Pompei

BBC Domesday / BBC Enterprise

“Kaderdatenspeicher” della Germania dell’Est / Archivi Federali tedeschi

City Archives of Antwerp / Sito web della città di Anversa; Progetto DAVID

Si ringrazia Mariella Guercio dell’Università di Urbino per la fondamentale collaborazione prestata per la riuscita di questo dossier.
--

\* Alessandra Ruggiero, docente di Lingua e cultura inglese nell’Università di Teramo, è stata membro del Comitato tecnico-scientifico, ed è tuttora consulente, del *Consorzio Glossa*.



## Prefazione

La Direzione Generale per i Beni Librari e gli Istituti Culturali ha voluto dedicare al tema della conservazione delle memorie digitali l'iniziativa di maggior rilievo per caratterizzare il proprio contributo al semestre di Presidenza italiana dell'Unione Europea, nella convinzione di affrontare un problema che non può essere confinato nell'ambito delle istituzioni culturali, ma che interessa l'intera società civile e politica.

Negli ultimi anni si è diffuso un atteggiamento eccessivamente fiducioso, quando non ingenuo, riguardo alla capacità delle moderne tecnologie informatiche e telematiche di risolvere problemi di comunicazione e conservazione del sapere e della conoscenza.

A tale convinzione spesso si accompagna la mancata consapevolezza della necessità, ormai improrogabile, di definire metodologie e prassi di lavoro che affianchino i processi di produzione del digitale al fine di ridurre al massimo il rischio della perdita della memoria.

A determinare questa situazione di diffusa emergenza intorno alla possibilità di garantire la memoria storica dell'elaborazione culturale odierna, concorrono due fattori in particolare. In primo luogo, l'atteggiamento superficiale ed ottimistico sulle capacità "autoregolative" dei processi industriali che governano la produzione del digitale, e in secondo luogo l'esiguità dei finanziamenti destinati (soprattutto per quanto riguarda le istituzioni culturali) a gestire tali processi.

L'indagine che presentiamo aveva l'ambizione di illustrare un numero maggiore di esempi, ma le difficoltà incontrate in corso d'opera sono state notevoli, pur avendo sin dall'inizio deciso di documentare solamente casi pubblici, ossia esperienze per le quali si è potuto accedere a documentazione di carattere ufficiale.

Eppure, non solo per quanto riguarda alcune situazioni verificatesi nel nostro paese, ma anche in circostanze di rilievo europeo e internazionale, quando gli estensori del rapporto hanno tentato di approfondire la materia attraverso interviste e colloqui diretti, spesso è stato difficile, se non impossibile, raccogliere informazioni sufficienti a portare a conclusione la redazione di una scheda rappresentativa del problema.

Riteniamo, con questo primo rapporto, di essere riusciti a definire solamente i contorni della punta di un iceberg. Altre significative esperienze di perdita totale o comunque consistente di memoria digitale rimangono nell'ombra, non solo perché nessuno ha ritenuto utile darne notizia, ma anche e soprattutto perché in molti casi è risultata del tutto assente la coscienza del problema ancor prima della difficoltà di rappresentarlo in termini adeguati.

Questo dossier costituisce, nelle intenzioni dell'ICCU, il primo di una serie di studi riguardanti la materia, e consente di delineare una panoramica piuttosto ampia. Infatti spazia oltre l'ambito europeo, con particolare riferimento a due significative esperienze statunitensi, abbraccia un arco temporale di oltre trent'anni, e infine, attraverso i casi esaminati, rende possibile evidenziare una molteplicità di fattori determinanti per il recupero, parziale o totale, delle memorie digitali.

Mi riferisco, in particolare, alle situazioni di recupero rese possibili dal rinvenimento, se non fortuito sicuramente fortunato, di attrezzature tecnologiche vetuste e obsolete, ma simili o uguali a quelle utilizzate per il processo di generazione dei data base, e sulla combinazione di tale circostanza con il ritrovamento di documentazione di supporto alla creazione delle banche dati e delle procedure.

E' da sottolineare che, paradossalmente, l'emergenza è stata superata in più di un caso grazie al recupero della "memoria orale", ossia attraverso l'individuazione delle persone che erano state protagoniste di quei processi, e che ricordavano ancora le specifiche di generazione dei file e degli archivi da recuperare.

In una società nella quale oltre il 90% dell'informazione viene prodotta direttamente in formato digitale che non può contare, come è avvenuto in una fase intermedia, su un corrispettivo tradizionale (cartaceo o non), è difficile ammettere che i processi produttivi che sono alla base della nascita e dell'esistenza dell'informazione debbano poggiarsi, per garantire alle generazioni future la trasmissione della memoria del presente, prevalentemente su fattori occasionali e fortuiti quali la memoria orale e l'archeologia industriale.

Luciano Scala  
Direttore ICCU

## Introduzione

Questo documento raccoglie sei casi di studio di singolare rilevanza nell'illustrazione delle problematiche relative alla conservazione a lungo termine delle memorie digitali. La storia di ciascuno dei sei archivi di cui si è rischiate la perdita, totale o parziale, contribuisce a disegnare il panorama di una Galassia Gutenberg in cui l'entusiasmo per le potenzialità della tecnologia, percepite come infinite, si accompagna alla sottovalutazione dei vincoli che l'uso di tale tecnologia impone perché il futuro del nuovo patrimonio digitale sia all'altezza delle aspettative.

Il percorso seguito per la scelta dei casi e l'elaborazione delle schede è stato lungo e complesso. La generosa disponibilità degli esperti italiani e stranieri contattati, che hanno fornito preziosi suggerimenti, ha contribuito all'individuazione di una rosa di progetti interessanti sui quali approfondire le ricerche. In molti casi, però, nonostante l'indubbio valore di talune esperienze è stato praticamente impossibile rintracciare le informazioni necessarie per dare al lettore un quadro coerente e completo della vicenda.

I sei archivi scelti, relativi a contesti storico-geografici e tecnologici differenti e riportati in questa raccolta in ordine cronologico, sono tutti significativi per le problematiche che sollevano in termini di longevità dei supporti fisici, obsolescenza tecnologica e necessità di migrazione verso sistemi e supporti più moderni. E le soluzioni adottate per scongiurare il rischio di una perdita totale dei dati raccolti evidenziano tutte la necessità di ricorrere a metodologie e risorse che spesso poco hanno a che vedere con la tecnologia in costante evoluzione.

I file statistici relativi al **Censimento USA del 1960** erano stati archiviati su nastri di tipo UNIVAC III-A e II-A o compatibili, e quando alla fine degli anni Settanta si è deciso di migrarli su supporti di memorizzazione più moderni le unità nastro UNIVAC II-A erano ormai obsolete, il che ha reso il progetto una vera e propria sfida tecnologica. L'esito positivo del recupero del 99,3% dei dati di microaggregazione è stato possibile grazie all'individuazione di strumentazioni uguali a quelle utilizzate al momento della creazione dell'archivio, mentre la percentuale di dati che non è stato possibile recuperare era archiviata su nastri andati perduti oppure deteriorati.

Il **Combat Air Activities File** (CACTA) contiene dati relativi alle missioni di attacco aereo USA in Vietnam (ottobre 1965 – dicembre 1970), comprese le coordinate geografiche dell'obiettivo di ciascun bombardamento, ed è stato creato con un particolare sistema di gestione delle banche dati detto "NIPS" sviluppato per il Dipartimento della Difesa USA. I dati contenuti nel file, dopo essere stati convertiti in un formato non proprietario codificato secondo lo standard EBCDIC, sono stati utilizzati dagli artificieri per localizzare sul territorio di Ho Chi Min le bombe inesplose che hanno

fatto dopo la fine della guerra migliaia di vittime; ma le coordinate geografiche così individuate erano sbagliate, poiché la migrazione ha introdotto nei record delle anomalie, che non sono state scoperte per quasi 25 anni e che è stato possibile correggere soltanto perché i National Archives avevano conservato i dati NIPS originali e li avevano periodicamente migrati su supporti più moderni in modo che rimanessero “vivi” o leggibili. Il che dimostra quanto sia importante conservare il flusso originale di bit.

La banca dati **Neapolis** relativa alle risorse ambientali e artistiche dell’area vesuviana, creata nel 1986 nell’ambito dei Giacimenti Culturali, costituisce il primo progetto nel settore dei Beni Culturali in cui sia stato realizzato un centro di documentazione integrato ed estremamente innovativo dal punto di vista delle tecnologie, basato su database relazionale, che raccoglie informazioni catalografiche (schede e immagini), cartografia e diari di scavo. Nel 1989, a progetto ultimato, l’elaboratore su cui risiedevano tutti i dati e le applicazioni è stato spento per la mancanza di risorse finanziarie e umane da parte della Soprintendenza Archeologica di Pompei, e quando nel 1999 si è deciso di verificare la possibilità di recuperarne il patrimonio di informazioni, si è dovuto prendere atto che riaccendere il sistema era un’operazione difficilmente realizzabile e che avrebbe richiesto un impegno economico insostenibile. Per tale motivo è stato avviato un progetto di recupero dei dati che è durato circa due anni e che è stato possibile solo per la disponibilità di un elaboratore simile a quello originario e per l’individuazione di persone che, avendo collaborato al Progetto Neapolis, con la loro memoria orale ne garantivano la fattibilità.

Il **BBC Domesday**, che ha costituito una pietra miliare nel campo dei prodotti multimediali, fu realizzato in collaborazione da BBC, Acorn, Philips e Logica per celebrare i 900 anni del *Domesday Book* originale e consiste in un database che descrive la Gran Bretagna come appariva agli occhi degli inglesi nel 1086. Il progetto si fondava su una complessa combinazione di hardware e software, di cui molti elementi non sono compatibili con i computer di oggi. Pochissimi utenti individuali e istituzionali dispongono ancora di sistemi BBC Domesday funzionanti, quindi il sistema si sta avviando rapidamente verso la completa obsolescenza. Elemento caratteristico del recupero – che rientra in un più ampio progetto sulle strategie di conservazione del patrimonio digitale (CAMiLEON) – oltre alla disponibilità di un vecchio BBC Domesday seppure semifunzionante, è stato lo sviluppo di un software di emulazione del sistema originario che dovrebbe costituire il futuro delle metodologie e tecniche di recupero di archivi a rischio.

Il **Kaderdatenspeicher** è il database che conteneva i file personali relativi ai membri dello staff di tutte le agenzie governative dell’ex-Germania dell’Est, realizzato su sistemi simili ai mainframe occidentali e archiviato su nastri spesso di cattiva qualità. Nell’ottobre del 1990, al momento dell’unificazione, fu necessario mettere al sicuro e preservare dalla possibile distruzione un gran

numero di archivi digitali, ma la documentazione che avrebbe dovuto accompagnare questi archivi elettronici (è il caso del *Kaderdatenspeicher*) risultò incompleta, se non addirittura assente, per cui è stato necessario prendere diverse misure per individuare e verificare la struttura dei file di dati e per ricostruire la documentazione mancante. L'esito positivo del recupero del *Kaderdatenspeicher* è stato reso possibile da due condizioni fondamentali: il reperimento di documentazione cartacea che, seppur non direttamente descrittiva dell'archivio, ha consentito l'individuazione della struttura dei file di dati e quindi la decifrazione del loro significato, e l'individuazione e la collaborazione di ex impiegati dei centri di elaborazione dati della Germania dell'Est che hanno contribuito, con la loro memoria orale, alla decifrazione di particolari algoritmi di compressione altrimenti incomprensibili. Infine, Anversa è stata la prima città belga ad avere, nel 1995, un sito web – *Digitale Metropool Antwerpen* – disponibile in quattro versioni (dal 1995 al 2001) archiviate su nastri di backup memorizzati in un formato proprietario e privi di documentazione di accompagnamento. Quando i nastri sono stati trasferiti agli Antwerp City Archives, nell'autunno del 2001, la possibilità di ricostruire le prime versioni di DMA erano molto incerte. L'esito positivo del recupero – parziale per l'indisponibilità di una delle quattro versioni dovuta al fatto che il nastro su cui era archiviata era stato sovrascritto – è stato reso possibile principalmente per il contributo di persone che ricordavano procedure non scritte e avevano familiarità con la vecchia configurazione. In sintesi, i sei casi scelti per descrivere i rischi che minacciano la possibilità di preservare a lungo termine le memorie digitali sottolineano che se, da un lato, la tecnologia permette a un numero sempre crescente di individui l'accesso alle informazioni prodotte su ampia scala, aumentando costantemente le possibilità di conoscenza, dall'altro tutto ciò richiede che la costituzione degli archivi digitali sia sempre accompagnata da documentazione adeguata che ne descriva le scelte e le procedure di creazione, e che si avvii una seria riflessione sulla definizione di standard che sono essenziali. Perché se il *medium* oggi è anche il messaggio, la sua inaffidabilità mina alle radici il futuro del nostro patrimonio.

Alessandra Ruggiero  
Università di Teramo



<b>Archivio</b>	Censimento USA del 1960
<b>Funzione</b>	File di riepilogo (“file di microaggregazione”) sul Censimento del 1960
<b>Ente responsabile</b>	Census Bureau / National Archives and Records Administration

### **Descrizione dell’archivio**

I risultati del Censimento USA del 1960 furono microfilmati nel 1961 per l’archiviazione a lungo termine. Sulla base delle informazioni raccolte in ciascun censimento, il Census Bureau crea, oltre ai microfilm, anche diversi rapporti di natura essenzialmente demografica che descrivono la composizione etnica della popolazione degli Stati Uniti, documentano l’andamento dei flussi migratori e arrivano a dire persino quante stanze da bagno ci sono nella casa dell’americano medio.

Nel 1961, lo staff del Census Bureau aveva accesso per la prima volta ai computer che, al fine di semplificare le modalità di analisi, vennero utilizzati per creare i “file di microaggregazione” che contengono informazioni statistiche. Nei precedenti censimenti queste informazioni erano state archiviate su schede perforate, ma negli anni sessanta erano i nastri magnetici il supporto di memorizzazione più all’avanguardia. I rapporti furono generati e stampati su carta; quindi i nastri furono conservati.

### **Rilevanza dell’archivio**

La Costituzione americana stabilisce che ogni dieci anni ci sia un censimento della popolazione, il cui scopo ufficiale è quello di ridistribuire i distretti per le elezioni del Congresso, ma che fornisce anche dati storici statistici relativi alla nazione e alla sua popolazione, oltre a costituire una risorsa economica e uno strumento di inestimabile valore. Il censimento, infatti, mostra non solo dove vive la popolazione, ma anche quali sono, ad esempio, i livelli di istruzione e di reddito.

Nel 1890, i dati vennero registrati con l’ausilio delle schede perforate di Hollerith, mentre nel 1950 venne utilizzato un computer UNIVAC per la tabulazione dei dati. Per il Censimento del 1960, il primo a utilizzare il servizio postale per la raccolta dei dati, il Census Bureau e il National Bureau of Standards hanno sviluppato il FOSDIC (Film Optical Sensing Device for Input to Computers), che è stato utilizzato fino al Censimento del 2000. I questionari, da completare contrassegnando la risposta appropriata, venivano fotografati su microfilm. Il FOSDIC leggeva quindi le risposte automaticamente e trasferiva i dati su nastro per farli poi immettere su computer.

### **Cosa ne è stato?**

In un memo tecnico interno del 1963, il Census Bureau elencava i file su nastro prodotti per il Censimento del 1960 e archiviati in modo “permanente”.

D’accordo con lo staff dei National Archives, nel 1975 il Census Bureau ha formulato un piano di conservazione per i dati del 1960. Sarebbero stati mantenuti 132 dei nastri compatibili e sarebbero stati copiati i file di 1.273 dei nastri di tipo III-A su nastri compatibili.

Nel 1975 e 1976, un membro della Machine Readable Archives Division dei National Archives ha riesaminato i file di microaggregazione o derivati che il Census Bureau aveva conservato su nastri di tipo II-A, individuando sette serie di microaggregazioni di livello basso da conservare a lungo termine. Queste sette serie si trovavano su 642 dei nastri di tipo II-A che il Census Bureau aveva deciso di migrare su supporti di memorizzazione più moderni, almeno moderni per gli standard industriali del 1975. A quel punto, però, le unità nastro erano obsolete, il che rese la conservazione di questi nastri una vera e propria sfida tecnologica.

### **Caratteristiche del recupero**

Nonostante i problemi, si riuscì a trovare alcune vecchie unità nastro ancora in uso che erano in grado di leggere i nastri. Tra l'altro, queste vecchie unità erano installate su un sistema che montava anche unità di memorizzazione più moderne, il che rese più semplice la conversione. Nel 1979, il Census Bureau era riuscito a copiare 640 dei 642 nastri di tipo II-A su nastri di tipo più recente. Gli altri due nastri, che non furono copiati perché erano in realtà scomparsi, contenevano 7.488 record, vale a dire circa lo 0,5 per cento del totale di circa 1,5 milioni di record memorizzati su nastri di tipo II-A che erano stati selezionati per l'archiviazione a lungo termine. Dei 640 nastri ritrovati, solo 1.575 record (vale a dire meno dello 0,2 per cento del numero totale di record preziosi memorizzati su nastri di tipo II-A) non poterono essere copiati perché deteriorati.

Di conseguenza, solo una piccola quantità di record relativi al Censimento del 1960 sono andati perduti, ed è successo a causa del controllo inadeguato dell'archivio e del deterioramento fisico di una piccolissima quantità di record. La cosa importante è che il 99,3 per cento dei dati di microaggregazione del 1960 sono stati salvati su nastri di tipo attuale e sono a tutt'oggi leggibili.

I risultati dell'analisi originale erano già stati pubblicati in formato cartaceo negli anni sessanta, e in quel formato sono ancora conservati. I documenti originali del Censimento del 1960 erano anche stati memorizzati su microfilm che sono ancora in buone condizioni, custoditi ai National Archives. In ottemperanza alle leggi americane, i documenti completi relativi al Censimento USA del 1960 conservati su microfilm saranno resi pubblici nel 2032, cioè 72 anni dopo la rilevazione originale.

### **Costo del recupero**

Non è stato possibile reperire informazioni sul costo complessivo delle attività di recupero.

### **Considerazioni finali**

L'esito positivo del recupero del 99,3% dei dati di microaggregazione è stato possibile per la disponibilità di strumentazioni uguali a quelle utilizzate al momento della creazione dell'archivio. I dati che non è stato possibile recuperare erano archiviati su nastri andati perduti oppure deteriorati.

<b>Archivio</b>	Combat Air Activities File (CACTA)
<b>Funzione</b>	Database sulle missioni di attacco aereo in Vietnam
<b>Ente responsabile</b>	National Archives and Records Administration

### **Descrizione dell'archivio**

Il Combat Air Activities File (CACTA) contiene dati relativi alle missioni di attacco aereo nell'Asia Sudorientale, che hanno coinvolto tutte le forze militari statunitensi e la forza aerea del Vietnam del Sud. Il "file" comprende 32 file di dati, ciascuno dei quali registra informazioni sulle missioni combattute in un arco di tempo di due mesi. I record descrivono le missioni aeree del periodo che va dall'ottobre 1965 al dicembre 1970. Non c'è traccia dell'esistenza di dati relativi al novembre 1967. Nel 1971, il file CACTA fu sostituito dal Southeast Asia Data Base (SEADAB). Ciascun record CACTA include: nome e data della missione; funzione e luogo della missione; tipo, numero e identificazione dell'aereo; esito della missione, inclusi dati su perdite e danni relativi all'aereo e all'equipaggio; commenti a testo libero. I record contengono anche dati sui bombardamenti in Vietnam, tra cui l'artiglieria utilizzata e le coordinate geografiche dell'obiettivo del bombardamento.

### *Tecnologie utilizzate*

Il Combat Air Activities File (CACTA), come molti dei file di dati relativi alla guerra del Vietnam custoditi presso il Center for Electronic Records dei National Archives, erano stati creati dal Dipartimento della Difesa utilizzando uno dei primi sistemi di gestione delle banche dati – il National Military Command System Information Processing System 360 Formatted File System, comunemente detto NIPS. La struttura dei dati dei file NIPS è gerarchica, nel senso che ciascun record di dati è costituito da dati fissi, che non si ripetono e che hanno un livello di dati subordinati. Ogni record è di lunghezza variabile ed è solitamente organizzato nei seguenti insiemi di elementi: un Control Set, nel quale si trova un identificativo univoco di record, come ad esempio il numero del rapporto sulle operazioni; un Fixed Set, che contiene dati non ripetitivi; e uno o più tipi di Periodic Set. Ogni tipo di Periodic Set può ricorrere una o più volte. Inoltre, i file NIPS possono contenere Variable Set che compaiono solo quando sono presenti dei dati.

### **Rilevanza dell'archivio**

Durante la Guerra in Vietnam, gli Stati Uniti hanno bombardato in modo massiccio delle parti del Sentiero di Ho Chi Min, nel Laos orientale – con più di 2 milioni di tonnellate di bombe – e ben il 30 % delle bombe sganciate non sono esplose. Più di trent'anni dopo, i coltivatori che lavoravano nella zona e i bambini che giocavano nelle distese di bambù hanno fatto esplodere casualmente le bombe, che hanno così ucciso o mutilato più di 10.000 persone.

Un programma umanitario finanziato a livello internazionale per individuare potenziali bombe inesplose dalla Guerra del Vietnam si basa su obiettivi, coordinate geografiche e tipi di artiglieria indicati nei due database sulle missioni aeree che riguardano gli anni 1965–1975, che i National Archives avevano acquisito nel 1976-77. Nel Laos i dati relativi ai bombardamenti sono stati inseriti in un sistema informativo territoriale che registra le coordinate dell'obiettivo nel database dei bombardamenti. Quindi sono state stampate delle mappe, utilizzate sul campo dagli artificieri che si servono di dispositivi GPS portatili per localizzare sul territorio le coordinate attraverso satelliti navigazionali.

### **Cosa ne è stato?**

Lo sforzo iniziale fu vano, poiché le coordinate geografiche erano sbagliate. La causa di tale errore stava nel fatto che i dati erano stati inizialmente creati con un sistema di generazione di report chiamato National Information Processing System (noto anche come NIPS). Il NIPS era stato

sviluppato per il Dipartimento della Difesa, ma quando il Dipartimento della Difesa ne aveva interrotto il supporto. Nel momento in cui gli archivi furono trasferiti ai National Archives lo staff della Sezione Archivi Digitali (Machine Readable Archives Division) iniziò un processo detto di “de-NIPSing” dei record, ovvero di esportazione dal sistema NIPS e riformattazione in un formato di file non proprietario codificato. I file “de-NIPS” così ottenuti non dipendono più dal software NIPS con cui erano stati creati e possono essere elaborati e manipolati usando applicazioni software ampiamente disponibili sul mercato.

### **Caratteristiche del recupero**

Per 25 anni si era ritenuto che i file “de-NIPS” fossero record riformattati affidabili. Invece, le anomalie riscontrate nel Combat Activities File hanno sollevato un problema. Pare che al momento dell’attività di “de-NIPSing” la documentazione che avrebbe dovuto accompagnare il file di dati fosse incompleta, se non addirittura mancante, poiché le coordinate geografiche, che erano codificate in binario (misurazione angolare binaria, una forma di “compressione dati”) per risparmiare spazio, erano state trattate erroneamente come ASCII a 7-bit in ciascun campo di dati. Di conseguenza, tutte le coordinate geografiche furono sbagliate. Una volta individuato il problema, fu avviato un processo per convertire correttamente questi campi.

### **Costo del recupero**

Non è stato possibile reperire informazioni sul costo complessivo delle attività di recupero.

### **Considerazioni finali**

L’osservazione importante da fare è che la migrazione ha introdotto nei record delle anomalie, che non sono state scoperte per quasi 25 anni e che è stato possibile correggere soltanto perché il NARA aveva conservato i dati NIPS originali (vale a dire gli 0 e 1 originali) e li aveva periodicamente migrati su supporti più moderni in modo che rimanessero “vivi” o leggibili. Il che dimostra quanto sia importante conservare il flusso originale di bit.

<b>Archivio</b>	Banca dati del <i>Consorzio Neapolis</i> - 1986
<b>Funzione</b>	Progetto di valorizzazione integrale delle risorse ambientali e artistiche dell'area vesuviana
<b>Ente responsabile</b>	Soprintendenza Archeologica di Pompei

### **Descrizione dell'archivio**

Nel corso del progetto, che ha avuto durata biennale (1987-1989) e un costo di 36 miliardi di lire (finanziamento ottenuto nell'ambito dei Giacimenti Culturali) e che è stato sviluppato da un team di lavoro che ha raggiunto circa 110 unità, sono state realizzate le seguenti attività:

- digitalizzazione della cartografia preesistente del territorio interessato, e in particolar modo della città antica di Pompei, e restituzione di rilievi aerofotogrammetrici
- censimento delle emergenze archeologiche diffuse sul territorio
- censimento dei monumenti archeologici del complesso di Pompei
- censimento dei beni architettonici, dei Centri Storici e dei beni ambientali del territorio
- catalogazione scientifica degli affreschi e dei mosaici rinvenuti nel complesso archeologico di Pompei
- acquisizione dei diari di scavo di Pompei (periodo 1862-1980) mediante schedatura ed immagini, con indicazione del periodo di competenza di ciascun documento e sua attuale collocazione
- schedatura di immagini d'epoca relative al complesso archeologico di Pompei e dei diari dei viaggiatori dei secoli XVII e XVIII.

### *Prodotti*

- Carta tecnica di base, in scala 1:5.000, del territorio, ottenuta da rilevamento aerofotogrammetrico e relativa restituzione e memorizzazione in forma digitale
- Diverse carte tematiche, in scala 1:25.000 (anch'esse memorizzate in forma numerica) e approfondimenti, in scala 1:5.000, per specifiche aree di interesse progettuale
- Realizzazione del *fotopiano*, in scala 1:500, ottenuto mediante voli ad hoc effettuati nel 1987, dell'area del complesso archeologico di Pompei, nonché rilievo aerofotogrammetrico e relativa restituzione numerica in scala 1:500 della stessa area. Sia il fotopiano che la pianta tecnica sono disponibili anche in formato digitale
- Planimetria in scala 1:1.000 della città antica di Pompei, articolata in *regiones* e *insulae*, ottenuta da una rivisitazione della carta di Van der Poel – grazie alla ricognizione fisica di circa venti archeologi che l'hanno corretta e integrata anche con i piani inferiori e superiori rispetto al piano terra – e dalle rilevazioni di tutte le strutture interne di interesse archeologico (forni, macine, impluvi, etc.)
- Indirizzario completo degli edifici di Pompei con informazioni sui nomi delle case (anche con i nomi che hanno avuto in vari periodi storici), sulle destinazioni d'uso e le tipologie di edifici che permette di analizzare le caratteristiche di Pompei sotto diversi aspetti: commerciale, di divertimento, di distribuzione delle strutture pubbliche e religiose
- Catalogazione scientifica degli affreschi e dei mosaici del complesso archeologico di Pompei secondo le norme indicate dall'ICCD, e relativo inserimento in una *banca dati*, attribuendo un codice di riferimento biunivoco sia alla facciata della parete (o del pavimento) che sostiene (o sosteneva prima del suo distacco) l'affresco (o il mosaico), sia alla scheda di catalogazione scientifica, nonché alla documentazione di corredo della scheda stessa, ottenendo così un rigoroso rapporto di relazione reperto-contesto fisico
- Realizzazione di programmi di gestione dei dati archiviati che consentivano, mediante l'uso di uno specifico linguaggio semplificato per l'utente, interrogazioni e correlazioni ricorrendo a ricerche sia di tipo relazionale che in forma libera

- Adozione, a corredo della scheda di catalogazione, di immagini numeriche a colori, per le loro caratteristiche di stabilità nel tempo rispetto a quelle fotografiche
- Acquisizione nella banca dati sia dei diari di scavo, sia delle immagini d'epoca (acquerelli, disegni) dei reperti
- Rappresentazione del contesto spaziale sia mediante le piante del monumento archeologico contenitore dei reperti, sia attraverso planimetrie dell'*insula* e delle *regiones* di appartenenza del monumento, nonché per mezzo di una planimetria dell'intero complesso archeologico.

Per l'intero progetto si è tenuto conto della riproducibilità dei dati archiviati, per poter diffondere i contenuti prodotti anche in altre sedi di studio e ricerca. I dati sono stati pertanto archiviati su dischi ottici e nastri magnetici.

#### *Tecnologie utilizzate*

Il sistema informativo realizzato operava su un elaboratore centrale di media potenza, che consentiva il funzionamento simultaneo e concorrente di diversi sottosistemi informatici. Il sistema operativo permetteva il funzionamento in multiutenza e ospitava un sistema relazionale, che gestiva la banca dati, unica e centralizzata, dell'intero sistema informativo.

Al sistema centrale erano collegate delle workstation, che svolgevano localmente alcune fasi elaborative sincrone al sistema centrale; tali workstation consentivano agli utenti sia di inserire nuovi dati (schede, immagini, mappe digitalizzate), sia di accedere coerentemente alle informazioni.

#### **Rilevanza dell'archivio**

Costituisce il primo progetto, nel settore dei Beni Culturali, in cui è stato realizzato un centro di documentazione integrato ed estremamente innovativo dal punto di vista delle tecnologie, basato su database relazionale, che raccoglie informazioni catalografiche (schede e immagini), cartografia e diari di scavo. Il sistema di navigazione integrata attraverso tutte le informazioni della banca dati fu il primo esempio, e tuttora rappresenta uno dei rari casi di banca dati dei beni culturali che permette una visione integrata di tutte le informazioni inerenti un determinato bene.

#### **Cosa ne è stato?**

Il sistema informativo realizzato nel corso del progetto è stato ufficialmente consegnato alla Soprintendenza Archeologica di Pompei, pienamente operativo, nel 1989. Nello stesso anno l'elaboratore su cui risiedevano tutti i dati e le applicazioni, collocato nel CED di Boscoreale, è stato spento per la mancanza di risorse finanziarie e umane, da parte del Ministero dei Beni Culturali e Ambientali.

Il sistema è stato riattivato, non senza difficoltà, nel solo periodo compreso tra l'ottobre 1989 e il maggio 1990, per consentire la realizzazione della mostra itinerante *Rediscovering Pompeii*, organizzata dal Ministero per i Beni Culturali e Ambientali.

Quando nel 1999 la Soprintendenza Archeologica di Pompei ha deciso di verificare la possibilità di recuperare il patrimonio di informazioni prodotto dal Progetto Neapolis, si è dovuto prendere atto che, la riaccensione del sistema, seppure fosse risultata possibile, avrebbe richiesto un impegno economico insostenibile.

Per tale motivo è stato avviato un progetto di recupero dei dati, concordato e approvato dall'AIPA, che ha coinvolto alcune delle professionalità che avevano contribuito allo sviluppo del progetto Neapolis. Il recupero è stato effettuato grazie alle risorse destinate dalla Soprintendenza e si è concluso nel 2000.

Ad oggi la Soprintendenza è impegnata in progetti di manutenzione, aggiornamento ed espansione della banca dati esistente.

### **Caratteristiche del recupero**

Il progetto di recupero ha riguardato prevalentemente la banca dati, che costituisce il patrimonio più consistente del progetto Neapolis. Quanto ai programmi per la gestione dei dati e per la fruizione, non è stato possibile recuperarli, dati i cambiamenti tecnologici avvenuti nel decennio di spegnimento del sistema. La società incaricata del recupero ha fornito alla Soprintendenza tre nuove applicazioni (per la catalogazione ICCD delle schede RA (reperti archeologici); per la gestione delle schede Diari di scavo; per la gestione della banca dati) e un software applicativo per la consultazione con tecniche Web delle informazioni integrate nella banca dati.

Quanto alla banca dati, al termine del progetto di recupero sono stati prodotti i seguenti materiali:

- 32.385 schede RA
- 21.732 schede Diari di scavo
- 432 schede CAT
- 337 schede BAT
- 7.093 immagini collegate alle schede RA
- 27.394 immagini dei Diari di scavo
- 29 immagini collegate alle schede CAT in formato JPG
- pianta archeologica 1:1.000 degli scavi di Pompei in formato ArcView
- pianta 1:25.000 delle emergenze archeologiche dell'area vesuviana in formato ArcView

In aggiunta al materiale prodotto sono stati consegnati alla Soprintendenza:

- 1 CD-ROM contenente:
  - backup delle mappe in formato ArcView e Autodesk Mapguide
  - backup del database SQLServer
  - copia, in formato Word, della documentazione cartacea consegnata
- 7 CD-ROM contenenti copia di tutte le immagini in formato JPG
- Documentazione cartacea:
  - Guida utente e Documentazione tecnica per l'installazione relative all'applicazione di consultazione della banca dati
  - Guida utente e Specifiche funzionali dell'applicazione "Gestione movimento Beni Culturali"

Condizione fondamentale per il recupero dei dati del Progetto Neapolis è stata la disponibilità, presso la società incaricata del recupero, di un elaboratore con caratteristiche simili a quello ormai in disuso di Boscoreale, completo di unità nastro, che ha permesso la lettura ed elaborazione di 140 nastri che costituivano buona parte dell'archivio di Neapolis, oltre che la ri-creazione dell'ambiente di lavoro.

### **Costo del recupero**

Il costo stimato dalle aziende incaricate del recupero è stato di circa 200.000 euro.

### **Conclusioni**

L'esito positivo del recupero della memoria del progetto Neapolis è stato reso possibile:

- 1) dalla disponibilità delle risorse economiche individuate dalla Soprintendenza
- 2) dalla disponibilità di un sistema di elaborazione con caratteristiche simili a quello utilizzato nel progetto
- 3) dalla individuazione e dalla disponibilità delle persone che avevano collaborato alla realizzazione del progetto conservandone la "memoria orale"

Il costo del recupero della banca dati è da considerarsi molto basso in rapporto all'entità dell'investimento originario.

<b>Archivio</b>	BBC Domesday
<b>Funzione</b>	Database che descrive la Gran Bretagna come appariva agli occhi degli inglesi nel 1986
<b>Ente responsabile</b>	BBC Enterprises

### **Descrizione dell'archivio**

Il Progetto BBC Domesday, che ha costituito una pietra miliare nel campo dei prodotti multimediali, fu realizzato per celebrare i 900 anni del *Domesday Book* originale – il documento, considerato la prima grande opera di burocrazia statale, commissionato da Guglielmo il Conquistatore per avere una sorta di “registro” della terra che aveva conquistato vent’anni prima, nel 1066.

Il BBC Domesday fu prodotto da BBC con un costo 2,5 milioni di sterline. I dati raccolti potevano essere visualizzati dagli utenti con una speciale configurazione hardware.

### *Prodotti sviluppati*

Il pacchetto relativo al Progetto Domesday conteneva due dischi: il Community Disc e il National Disc. La differenza fondamentale tra i due sta nel fatto che nel Community Disc le informazioni sono organizzate su base geografica e mostrano la Gran Bretagna come appariva agli occhi di chi ci viveva, mentre nel National Disc le informazioni sono organizzate per argomenti e danno una panoramica generale della nazione.

Il Community Disc si apre con una mappa “navigabile” del Regno Unito dalla quale si possono visualizzare le informazioni – foto e testi – che gli studenti hanno fornito sul loro quartiere.

Il National Disc raccoglie un’ampia varietà di dati statistici, immagini del Regno Unito e della sua cultura in una serie di fotografie, passeggiate virtuali (un primo esempio di realtà virtuale) e una gran quantità di video.

### *Tecnologie utilizzate*

BBC Domesday è eseguibile su computer BBC Master modificati con l’aggiunta di diversi componenti hardware speciali. Il BBC Master era un computer a 8 bit molto diffuso, insieme al “BBC Modello B”, in Gran Bretagna e in Australia a metà degli anni Ottanta. Al Master furono aggiunti una scheda SCSI, un coprocessore, una trackball e un tipo particolare di lettore di videodischi detto LVROM.

L’applicazione multimediale del BBC Domesday fu scritta in un linguaggio di programmazione detto BCPL, che era stato scelto per garantire una certa portabilità su diverse piattaforme ma che, sebbene abbia influenzato lo sviluppo di linguaggi di programmazione come il C e Java, in realtà non mai è stato utilizzato granché.

Il software del Domesday fornisce un’interfaccia multimediale per visualizzare la gran quantità di dati memorizzata sui videodischi. Le funzioni di ricerca avanzata, l’interfaccia di navigazione intuitiva e la possibilità di contenere diversi tipi di informazione in un’unica schermata fecero del Domesday un progetto all’avanguardia.

### **Rilevanza dell'archivio**

Il BBC Domesday è stato uno dei più importanti prodotti interattivi del suo tempo ed è stato realizzato su un’ampia scala, come mai si era visto fino ad allora. I due dischi contenevano più di 250.000 luoghi geografici, 25.000 mappe, 50.000 immagini, 3.000 serie di dati, 60 minuti di video e un numero imprecisato di parole. Più di un milione di persone, tra studenti e ricercatori provenienti

da tutto il paese, hanno preso parte a questo ambizioso progetto, e si dice che ci vorrebbero più di sette anni per guardare l'intero contenuto dei dischi. Sebbene sia stato un grande insuccesso commerciale, chi lo ha realizzato lo ha ritenuto un trionfo per quanto concerne la progettazione interattiva.

### **Cosa ne è stato?**

Nel novembre 1986 i dischi vennero pubblicati dalla BBC Enterprises. Venivano venduti insieme all'hardware necessario per circa 5.000 sterline, più del doppio di quanto si era stimato potessero costare all'inizio del progetto e un po' troppo costosi per molte scuole che avrebbero dovuto acquistarli. A quell'epoca il Governo inglese aveva cambiato politica, decidendo di sovvenzionare le scuole all'acquisto del software piuttosto che dell'hardware. Gli sforzi delle società produttrici e persino l'aiuto dei finanziamenti comunitari ESPRIT non bastarono a ridurre sufficientemente i costi. Di conseguenza, furono pochi i computer ad essere acquistati e ne è sopravvissuto solo un piccolissimo numero.

Il BBC Domesday si fonda su una complessa combinazione di hardware e software che era stata progettata e realizzata nel 1986. Molti elementi di questa configurazione originale non sono compatibili con i computer di oggi. Pochissimi utenti individuali e istituzionali dispongono ancora di sistemi BBC Domesday funzionanti, che peraltro stanno diventando rari. Sebbene il BBC Master sia un computer incredibilmente resistente, dopo oltre 15 anni d'uso molti dei lettori hanno ormai raggiunto la fine del loro ciclo di vita. In breve, il sistema BBC Domesday si sta avviando rapidamente verso la completa obsolescenza.

Paradossalmente, la versione digitale, di soli 15 anni, è diventata illeggibile mentre la vecchia versione cartacea è custodita in buone condizioni al Public Record Office di Kew ed è ancora perfettamente utilizzabile.

### **Caratteristiche del recupero**

Il progetto CAMiLEON – Creative Archiving at Michigan and Leeds Emulating the Old on the New – ha sviluppato per tre anni delle strategie di conservazione del patrimonio digitale e ha utilizzato il BBC Domesday per verificare l'affidabilità delle strategie di emulazione sviluppate.

Il team CAMiLEON ha avuto accesso a un sistema Domesday semi-funzionante donato dalla School of Geography dell'Università di Leeds. Una delle prime attività effettuate per conservare il BBC Domesday è stata quella di trasferire i file di dati dai dischi laser da 12 pollici in flussi di byte accessibili su moderni dispositivi hardware. Un PC Linux è stato collegato al lettore di dischi laser con un cavo SCSI, in modo da poter leggere su PC gli articoli di testo e il database. Le immagini, compresi i singoli fotogrammi dei video, sono state trasferite su PC utilizzando una scheda video standard per catturare le immagini alla massima risoluzione, e poi sono state memorizzate in formato non compresso per evitare qualunque perdita di qualità. In totale, sono stati trasferiti circa 70 GB di dati per lato di ciascun disco laser.

Il passo successivo è stato quello di sviluppare un software che emulasse il computer BBC Master modificato e il lettore di dischi laser su cui girava il sistema BBC Domesday originale. Il team del CAMiLEON è riuscito ad emulare in modo eccellente il BBC Domesday con il prototipo "DomesEm", che comprendeva il coprocessore, le comunicazioni SCSI e le varie funzioni del lettore di dischi laser. A partire dall'emulatore BBC Micro "BeebEm" sviluppato da David Gilbert e Richard Gellman, CAMiLEON ha sviluppato il software necessario per supportare le funzionalità aggiuntive richieste per eseguire il BBC Domesday. I risultati di questo progetto di emulazione sono stati presentati a un pubblico selezionato di esperti in conservazione digitale e membri del team che aveva sviluppato il BBC Domesday originale il 2 dicembre 2002.

### **Costo del recupero**

Non è stato possibile reperire informazioni sul costo complessivo delle attività di recupero. Va comunque tenuto conto del fatto che il recupero rientra in un più ampio progetto sulle strategie di conservazione del patrimonio digitale.

### **Considerazioni finali**

Elemento caratteristico del recupero – oltre alla disponibilità di un vecchio sistema Domesday, seppure semifunzionante – è stato lo sviluppo di un software di emulazione del sistema originario che dovrebbe costituire il futuro delle metodologie e tecniche di recupero di archivi a rischio.

<b>Archivio</b>	“ <i>Kaderdatenspeicher</i> ” della Germania dell’Est
<b>Funzione</b>	Database di file personali dei funzionari di partito della Germania dell’Est
<b>Ente responsabile</b>	Archivi Federali Tedeschi ( <i>Bundesarchiv</i> )

### **Descrizione dell’archivio**

Il *Kaderdatenspeicher* contiene dati personali su 331.980 membri dello staff di tutte le agenzie governative dell’ex-Germania dell’Est (al 1989), tranne quelle del Ministero per la Sicurezza Nazionale, del Ministero della Difesa e del Ministero dell’Interno.

#### *Tecnologie utilizzate*

In generale, e anche nel caso del *Kaderdatenspeicher*, i centri informatici della Germania dell’Est utilizzavano sistemi mainframe per elaborare grandi quantità di dati. Il patrimonio di dati aveva solitamente una struttura gerarchica non troppo complicata. L’hardware e il software utilizzati erano copie e varianti di modelli occidentali, che ovviamente avevano nomi differenti. Ad esempio, i cosiddetti sistemi mainframe ESER della Germania dell’Est erano copie di mainframe a larga diffusione

Come supporti di memorizzazione, venivano utilizzati prevalentemente nastri a 9 tracce, molti dei quali avevano una densità di soli 800 bpi. Per problemi di produzione, questi nastri si dimostrarono di cattiva qualità. Era necessario rimuovere colla o abrasioni dai nastri prima di poterli leggere, e a volte gli strati del nastro si separavano dopo il primo utilizzo.

L’etichettatura dei nastri seguiva gli standard internazionali quasi senza alcuna variazione. Come per le applicazioni occidentali per mainframe, il codice usato era EBCDIC, insieme a quello russo DKOI (chiamato anche “ESER-Code” nell’ex Germania dell’Est), che tradotto significa Codice Binario per lo Scambio di Informazioni. Il codice DKOI è molto simile a quello EBCDIC e ne costituisce sostanzialmente un ampliamento con alcune variazioni.

### **Rilevanza dell’archivio**

I file del *Kaderdatenspeicher* non forniscono semplicemente dettagli sulla carriera politica e professionale degli ufficiali, ma anche informazioni sui loro genitori. Esistevano diverse copie del *Kaderdatenspeicher*, l’unica delle quali ancora esistente è quella acquisita dagli Archivi Federali. Almeno in un caso, ci sono prove sufficienti che una copia del *Kaderdatenspeicher* sia stata deliberatamente cancellata poco dopo l’unificazione tedesca allo scopo di proteggere i membri del partito. Il valore considerevole di questo patrimonio ha fatto sì che gli Archivi Federali decidessero di fare un investimento consistente per ricostruirne la documentazione.

### **Cosa ne è stato?**

Nell’ottobre del 1990, dopo 45 anni di separazione e di tradizioni istituzionali differenti, gli Archivi Centrali di Stato della ex-Germania dell’Est si sono fusi con quelli della Germania dell’Ovest. Quando le agenzie governative e le istituzioni della ex-Germania dell’Est furono improvvisamente chiuse, non solo i loro archivi cartacei, ma anche un gran numero di archivi digitali dovettero essere messi al sicuro o preservati dalla possibile distruzione. Molto spesso, però, come nel caso del *Kaderdatenspeicher*, la documentazione che avrebbe dovuto accompagnare questi archivi elettronici risultò incompleta, se non addirittura assente, per cui è stato necessario prendere diverse misure per individuare e verificare la struttura dei file di dati e per ricostruire la documentazione mancante.

Dopo aver copiato i nastri del *Kaderdatenspeicher*, sono state stampate le etichette, le intestazioni e i primi blocchi di dati di ciascun file. Le etichette e le intestazioni seguivano schemi già conosciuti, per cui è stato facile decifrarle. Da questi dati è stato possibile ottenere informazioni sul contenuto

di ciascun nastro e farsi una prima idea delle diverse generazioni e applicazioni del *Kaderdatenspeicher*.

In ogni caso, un problema che poi sarebbe risultato frequente è comparso già in questa fase iniziale. Nelle poche righe dell'etichetta e delle intestazioni erano stati utilizzati tre modi diversi per esprimere la data. Inoltre, i dati contenuti nel *Kaderdatenspeicher* rivelavano che solo il nome completo, il Codice di Identificazione Personale, l'indirizzo e l'agenzia erano scritti in EBCDIC. Tutti gli altri campi erano codificati con valori numerici rappresentati sotto forma di numeri binari.

### **Caratteristiche del recupero**

Senza una descrizione precisa della struttura dei file di dati, non c'era alcuna possibilità di comprendere il significato dei dati. Perciò fu subito chiaro che era necessario recuperare quante più informazioni possibili sul *Kaderdatenspeicher*. Il *Kaderdatenspeicher* fu creato ed elaborato su ordine del Consiglio dei Ministri, che scelse anche le commissioni. I *vertical file* di questo ufficio erano stati inseriti nelle collezioni degli Archivi Federali a Potsdam dopo l'unificazione. Dopo aver ricercato in questi archivi riferimenti al progetto del *Kaderdatenspeicher*, furono trovati una serie di record che contengono descrizioni del *Kaderdatenspeicher* e rapporti preparati dal centro elaborazione dati con molte informazioni importanti.

Questi documenti cartacei hanno fornito informazioni sul contenuto, lo scopo, la storia e lo sviluppo del progetto *Kaderdatenspeicher*, e in particolare:

- chi ha progettato il *Kaderdatenspeicher* e chi ha dato gli ordini,
- quali agenzie hanno collaborato,
- quali sono stati i diversi obiettivi del *Kaderdatenspeicher*,
- quali informazioni erano contenute nel *Kaderdatenspeicher*,
- come sono state raccolte le informazioni,
- quali versioni e aggiornamenti del *Kaderdatenspeicher* esistevano e quali centri informatici le avevano elaborate e archiviate,
- chi aveva accesso alle singole porzioni di informazioni contenute nel *Kaderdatenspeicher*,
- come sono state utilizzate le informazioni.

I rapporti al Consiglio dei Ministri contenevano anche informazioni sulla struttura dei file di dati e sui volumi di codice. Il *Kaderdatenspeicher* è formato da compilazioni annuali, le cosiddette "generazioni" di file di dati per il 1980 e per tutti gli anni dal 1985 al 1989, e da estratti preparati per vari scopi. Quasi tutti questi file di dati hanno una struttura almeno leggermente differente. Ciononostante, è stato possibile ricostruire la struttura dei file di dati di tutte le generazioni del *Kaderdatenspeicher*. Gran parte delle informazioni è stata dedotta dalle cosiddette "tabelle di indirizzo", che rappresentano il layout dei record di un determinato file.

I file di dati del *Kaderdatenspeicher* erano strettamente collegati ai cosiddetti database dello staff di ministeri e uffici governativi, che contenevano dati personali riguardanti i membri dello staff. Tutti i dati del *Kaderdatenspeicher* erano stati originariamente tratti da questi database dello staff. Gli Archivi Federali sono riusciti ad acquisire documentazione sufficientemente completa ed esaustiva sul database dello staff del Ministero della Pubblica Istruzione. Di conseguenza, è stato possibile desumere informazioni aggiuntive sul layout dei record e sulla struttura dei file di dati del *Kaderdatenspeicher* dalla documentazione del database dello staff di questo ministero.

Allo scopo di ottenere informazioni di riferimento, gli archivisti hanno anche contattato ex impiegati dei centri di elaborazione dati della Germania dell'Est che avevano creato gli archivi che erano stati acquisiti, o che ci avevano comunque lavorato. In rari e difficili casi, ad esempio quando erano stati usati algoritmi di compressione che non si riusciva a decifrare, i programmatori dei centri informatici della ex-Germania dell'Est furono addirittura assunti come consulenti.

### **Costo del recupero**

Il costo del recupero, effettuato grazie a uno sforzo congiunto degli Archivi Federali, dell'Università di Jena e del Center for Historical Social Research di Colonia, è stato stimato dalla direzione dei *Bundesarchiv* in una cifra che va dai 150.000 ai 200.000 euro, due terzi della quale a carico degli Archivi Federali.

### **Considerazioni finali**

L'esito positivo del recupero del *Kaderdatenspeicher* è stato reso possibile da due condizioni fondamentali: il reperimento di documentazione cartacea che, seppur non direttamente descrittiva dell'archivio, ha consentito l'individuazione della struttura dei file di dati e quindi la decifrazione del loro significato, e l'individuazione e la collaborazione di ex impiegati dei centri di elaborazione dati della Germania dell'Est che hanno contribuito, con la loro memoria orale, alla decifrazione di particolari algoritmi di compressione altrimenti incomprensibili.

<b>Archivio</b>	Sito web della città di Anversa
<b>Funzione</b>	Conservazione dei vecchi siti web della città di Anversa (4 versioni: 1995-2001)
<b>Ente responsabile</b>	City Archives of Antwerp; Progetto DAVID

### **Descrizione dell'archivio**

Nel 1995 Anversa ha avuto il suo primo sito web, che si chiamava DMA, conteneva informazioni generali sulla città (organi amministrativi, enti, eventi, agenda, news, siti web privati dei cittadini, etc.) ed era stato sviluppato da un nutrito team di web designer, webmaster di tutti gli uffici amministrativi e volontari. Gli utenti avevano diverse funzioni a disposizione: ricerca, link ad altri siti web, software scaricabile per web developer, chat café, giochi, newsgroup, etc.

### *Tecnologie utilizzate*

HTML, immagini GIF e JPEG, script CGI, Java script, Apache webserver.

### **Rilevanza dell'archivio**

Anversa è stata la prima città belga ad avere un sito web, denominato *Digitale Metropool Antwerpen* (DMA). I primi siti di città americane ed europee come ad esempio il *Digitale Stad Amsterdam* sono serviti come fonte di ispirazione per lo sviluppo di DMA. La versione 1.0 (grafica e non) è stata messa online l'11 giugno 1995. Poco dopo, a metà dicembre 1995, era stata già lanciata la versione 2.0 (con frame e senza frame). Entrambe le versioni hanno un valore storico che giustifica la loro conservazione.

### **Cosa ne è stato?**

I siti web erano stati archiviati su nastri di backup dallo staff di informatici dell'amministrazione cittadina e trasferiti agli Antwerp City Archives nell'autunno del 2001.

Telepolis, però, il centro informatico della città di Anversa, non possedeva la configurazione hardware e software necessaria per ripristinare su un hard disk i dati contenuti sui nastri. I backup erano stati effettuati in un formato proprietario su un computer SPARC con Solaris e una vecchia versione di Unix come sistema operativo. Telepolis non utilizzava più né quel tipo di computer, né quel software di backup. Inoltre, i file erano stati memorizzati su nastro con il sistema di registrazione del computer originario - e non vi erano computer del genere in uso - e non c'era documentazione disponibile. Pertanto, la possibilità di ricostruire le prime versioni di DMA erano molto incerte. In particolare, c'erano problemi con la vecchia configurazione di backup delle versioni 1.5 e 2 (non più disponibile), e con il nastro di backup della versione 3, che era stato sovrascritto.

### **Caratteristiche del recupero**

Il centro informatico è riuscito a ricreare la vecchia configurazione di backup e a ripristinare il contenuto dei nastri di backup obsoleti su CD-ROM. Dal momento che non c'era documentazione scritta disponibile, il motivo principale che ha reso il recupero possibile è stato l'aiuto di persone che ricordavano procedure non scritte, password e vecchi linguaggi di programmazione, e di persone che avevano familiarità con la vecchia configurazione. Inoltre, il ripristino delle versioni 1.5, 2 e 4 fu possibile perché i nastri contenevano ancora i dati originali, mentre quella della versione 3 non fu possibile, perché i dati originari erano stati sovrascritti.

I CD furono quindi trasferiti agli archivi cittadini e verificati. In generale, i file erano al 99 % indipendenti dalla piattaforma utilizzata. L'unica eccezione erano i link interni, assoluti, e il contatore, che era basato su script CGI. Il contatore non costituiva un problema, dal momento che non sarebbe stato attivo nella versione d'archivio. I link, invece, dovevano essere utilizzati anche

nelle versioni recuperate, il che generava qualche problema. Per molti link interni erano stati utilizzati percorsi assoluti, per cui quando si navigava nel sito web offline su CD, comparivano molti messaggi di errore oppure venivano visualizzate delle caselle con una croce rossa al posto di un'immagine mancante. Il numero di link perduti fu verificato con l'ausilio di "Linkbot", un programma di controllo automatico. La verifica della prima versione riportò come esito 2.509 link perduti, mentre il sito web aveva in totale più di 64.000 link. Il programma era in grado di indicare la pagina web che conteneva il link perduto e il singolo link che non faceva riferimento a un file preciso. La correzione vera e propria, però, doveva essere fatta a mano. Inoltre, i file scaricabili di formato obsoleto vennero migrati in formati più recenti o più adatti per l'archiviazione. Ad esempio, i testi furono trasformati in file XML. Dopo la conversione, i link a questi documenti sono stati aggiornati.

### **Costo del recupero**

Il costo del recupero è stato stimato dalla direzione degli Archivi municipali della città di Anversa in tre mesi di attività, ripartiti come segue.

- Personale informatico: 8 settimane di lavoro (3 settimane per ricreare la vecchia configurazione di backup; 4 settimane per aggiornare i link interni; 1 settimana per creare i CD, le versioni archiviabili, e per raccogliere i metadati)
- Archivista: 4 settimane di lavoro per sviluppare la procedura, aggiornare i link, catturare le schermate, e raccogliere i metadati.

### **Considerazioni finali**

L'esito positivo del recupero – parziale per l'indisponibilità di una delle quattro versioni dovuta al fatto che il nastro su cui era archiviata era stato sovrascritto – è stato reso possibile principalmente per il contributo di persone che ricordavano procedure non scritte e avevano familiarità con la vecchia configurazione.

