

# 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 "autoregolatrici" 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)
- Archivist: 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.



# Preservation of digital memories: risks and emergencies

## Six case studies

Edited by  
Alessandra Ruggiero \*

### *INDEX*

PREFACE by Luciano Scala

INTRODUCTION by Alessandra Ruggiero

#### THE CASE STUDIES:

US 1960 Census. Census Bureau / National Archives and Records Administration

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

Database of the *Consorzio Neapolis* / Soprintendenza Archeologica di Pompei

BBC Domesday / BBC Entreprise

East Germany “Kaderdatenspeicher” / German Federal Archives

City Archives of Antwerp / Web site of the City of Anversa; DAVID Project

<p>We thank Mariella Guercio (University of Urbino) for her important collaboration for the realization of this dossier.</p>
--

\* Alessandra Ruggiero, professor of English Language and Culture at the University of Teramo, has been a member of the Scientific Committee of the *Consorzio Glossa* and is still an advisor of the *Consorzio*.



## Preface

The General Direction for Library Heritage and Cultural Institutes dedicated to the subject of digital memory preservation the most relevant initiative in order to characterize its own contribution to the European Union Italian Presidency semester, in the conviction of facing a problem that can't be restricted to cultural institutions, but that affects the entire society.

In the past few years, an excessively faithful, when not ingenuous, attitude has spread about the possibilities of modern information and communication technology of solving the most complex problems arising from the communication and the preservation of knowledge.

This conviction is often accompanied by the lacking of awareness that we cannot delay any longer the definition of the methodologies and practices to lead the processes of digital resources production, in order to limit as much as possible the risks of losing digital memory.

Two factors in particular contribute to this situation of widespread emergency about the possibility of ensuring the historical memory of the contemporary culture for future generations. First, the superficial and optimistic attitude about the capability of industrial processes governing the digital memory production to regulate themselves, secondly the quite poor financial support (especially for what concerns cultural institutions) to manage such processes.

The survey we present had the ambition to illustrate much more examples, but we found heavy difficulties carrying out this work, even though we decided from the beginning to limit ourselves to experiences for which official documentation was available.

Yet, when the report's editors tried to go into more details, by means of interviews and conversations, it was often hard, if not impossible, to gather enough information to create a comprehensive record of the problem, and not only working on some situations in our country, but even in some situations of European and International relevance.

We think that, with this first report, we only drew the contour of an iceberg's top. Other important experiences of total or relevant loss of digital memory still remain in the shadow, not only because nobody thought it could be useful to report them but also, and above anything else, because, in many cases, the awareness of the problem turned out to be completely lacking, even before the difficulty to find out the suitable terms to define the problem itself.

This dossier is, in our intention, the first of a series of studies about this subject, and it allows drawing a rather broad picture. In fact it goes beyond the European context, with a special reference to two relevant experiences in the US. Furthermore, it covers a 30 years long period. Finally,

through the cases considered, it makes possible to point out several factors, which are relevant for the partial or total recovery of digital memory.

I'm referring, in particular, to those recovery experiences which were made possible both by the discovery, certainly lucky if not fortuitous, of obsolete technological equipment similar or equal to those used for building databases, and by the discovery of documentation about the creation of databases and the related procedures.

It must be pointed out that, strangely enough, the emergency was overcome in more than one case thanks to the *oral memory* recovery, that is, individuating people who were protagonists of those processes and who still remembered the specification for the creation of the files to be recovered.

Since over the 90% of information is currently produced in a digital format, without equivalent in a traditional format, we find difficult to admit that the information production processes, on which the origin and the existence of information are based, must mainly count, to guarantee to future generation the transmission of the present memory, on accidental and fortuitous factors like oral memory and industrial archaeology.

Luciano Scala  
Director of ICCU

## Introduction

The current document is a collection of six case studies notably important in illustrating the problems relating to the long term conservation of digital archives. Each of the six archives described has at some point risked total or partial loss, and their stories help to illuminate the panorama of a Gutenberg Galaxy in which enthusiasm for the vast potential of technology, perceived as being infinite, has been accompanied by the dramatic undervaluing of the limits which the use of such technology brings about, in order to have the future of this new digital heritage meet its great expectations.

The path followed for the selection of the case studies and the elaboration of their details is both long and complex. Experts from both Italy and other countries have been generous and helpful in offering valuable insights which have been useful in drawing up a shortlist of interesting projects into which deeper research has then been carried out. In many cases, however, despite the indubitable value of such experiences it has been practically impossible to trace all the information necessary to give the reader a coherent and complete overall picture of the affair.

The six archives selected, relative to specific historical-geographical contexts, are all significant due to the problems which arise in terms of longevity of physical supports, technological obsolescence and the necessity of migrating to newer and more modern systems and supports. The solutions adopted in order to avoid the risk of a total loss of the data gathered, moreover, highlight the vital necessity of going back to methods and resources which often have little to do with the constantly evolving technology.

The statistical files relative to the **1960 US Census** were stored on UNIVAC III-A and II-A tapes, or similar compatible ones. When it was decided to transfer them onto more modern supports at the end of the 1970's the UNIVAC II-A tape units were already obsolete, thus rendering the project a genuine technological challenge. The successful recovery of 99.3% of the micro-aggregation data was made possible due to the identification of hardware similar to that used when the archives were originally created, while the percentage of data which it was not possible to recover was stored on tapes that had been lost or had deteriorated.

The **Combat Air Activities File** (CACTA) contains data relating to the US air attacks in Vietnam (October 1965 – December 1970), including the geographic co-ordinates of the objective of each attack, and was created with a particular database management system called “NIPS” developed for the US Defense Department. The data stored on the files, after having been converted into a non-proprietary format coded following EBCDIC standards, was used by bomb technicians to identify

the parts of the Ho Chi Min Trail where unexploded bombs had caused thousands of victims following the end of the war. However, the geographic co-ordinates identified in this way turned out to be mistaken as the data transfer had created a number of anomalies in the records which were not discovered for 25 years. It was only possible to correct them because the National Archives had conserved the original NIPS data and had periodically transferred them onto more modern supports so that they could remain “live”, or legible. This shows just how important it is to preserve the original flow of bits.

The **Neapolis** database containing information on the environmental and artistic resources in the area around Mount Vesuvius, created in 1986 with financing obtained as part of the *Giacimenti Culturali* project, was the first cultural heritage project in which there was (from a technological point of view) an integrated and highly innovative documentation centre, based on a relational database which gathered catalogue information (files and images), cartography and excavation diaries. In 1989, when the project was completed, the mainframe system in which all the data and applications were held was switched off due to the *Ministero dei Beni Culturali*'s lack of financial and human resources. When in 1999 it was decided to investigate the possibility of recovering the wealth of information there, it was necessary to take into account that re-starting the system would be an operation almost impossible to achieve and which would have required an impractical economic commitment. For this reason, a project of data recovery was initiated which lasted for some two years and was made possible only due to the availability of a mainframe system similar to the original one and the identification of the people who had taken part in the original Neapolis project and were able to guarantee the practicability of the project with their oral memory. The **BBC Domesday**, a milestone in the field of multimedia production, was produced as a collaboration between BBC, Acorn, Philips and Logica to celebrate the 900th anniversary of the original *Domesday Book*. It consisted of a database which described Great Britain as it appeared through the eyes of the British in 1086. The project was created with a complex combination of hardware and software, many elements of which were no longer compatible with today's computers. Very few individual and institutional users today still have functioning BBC Domesday systems so the system is rapidly heading towards complete obsolescence. A typical element of the recovery – which is part of a larger project on the strategies of digital heritage preservation (CAMiLEON) – as well as the availability of an old BBC Domesday system (albeit a semi-functional one), has been the development of a software which emulated the original system and which could make up the future of methodologies and techniques of recovery of endangered archives.

The **Kaderdatenspeicher** is the database which contained the personal files on staff members of all the governing bodies of the former East Germany, created on systems similar to western mainframes and stored on tapes which were often of poor quality. In October 1990, with the unification, it was necessary to make a large number of digital archives safe and protect them from possible destruction, yet the documentation which should have accompanied these electronic archives (as in the case of the *Kaderdatenspeicher*) turned out to be incomplete, if not completely absent, and for this reason it was necessary to take a number of different measures to identify and verify the structure of the data files and to recreate the missing documentation. The positive outcome of the *Kaderdatenspeicher* recovery was made possible due to two fundamental conditions: the discovery of paper-based documentation which, even though not directly pertaining to the archive, made it possible to identify the structure of the data files and therefore the deciphering of their meaning, and the identification and collaboration of former employees of data elaboration centres in East Germany who contributed, with their oral memory, to the deciphering of certain otherwise incomprehensible compression algorithms.

Finally, in 1995 became Antwerp the first Belgian city to have a website - *Digitale Metropool Antwerpen* – available in four versions (from 1995 to 2001) and archived on backup tapes in a proprietary format and without support documentation. When the tapes were transferred to the Antwerp City Archives in autumn 2001, the possibility of reconstructing the first versions of *DMA* seemed to be highly uncertain. The positive outcome of the recovery – only partial due to the unavailability of one of the four versions due to the fact that the tape on which it had been stored had been overwritten – was made possible largely due to the contribution of people who remembered non-written procedures and were familiar with the old configurations.

To sum up, the six cases chosen to describe the risks which threaten the possibility of preserving digital records on a long-term basis underline that if, on one hand, technology permits an ever-increasing number of individuals to access information produced on a large scale, continually enlarging the possibilities of knowledge, on the other hand all this means that the creation of digital archives must always be accompanied by adequate documentation which describes the choices and the procedures involved in their creation, and which necessitates a serious reflection on the definition of essential standards. If today the *medium* is also the message, its undependability could fundamentally undermine the future of our heritage.

Alessandra Ruggiero  
University of Teramo



<b>Archive</b>	US 1960 Census
<b>Purpose</b>	Summary tape files (“micro-aggregation files”) about the 1960 Census
<b>Institution in charge</b>	Census Bureau / National Archives and Records Administration

### **Contents**

The US 1960 Census returns were microfilmed in 1961 for long-term storage. In addition to the microfilm, the Census Bureau also creates many reports from the information obtained in each census. These reports are mostly demographic in nature. They describe the ethnic make-up of the U.S. population; they document American migration patterns and even tell how many bathrooms are in the average American home.

In 1961, the staff at the Bureau of the Census had access to computers for the first time. In order to simplify some of the data analysis that the Census Bureau must conduct, they used the new computers to create the “micro-aggregation files” that contain statistical information. This information had been entered on punch cards in earlier censuses, but magnetic tape was the storage medium of choice in the '60s. The Bureau of the Census had the required data stored on 9,121 reels of magnetic tape: 7,297 reels “readable” with UNIVAC II-A tape drives; 1,678 tapes “readable” with UNIVAC III-A tape drives, and another 146 magnetic tapes created on still other brands of tape drives and “readable” with then contemporary industry-compatible tape drives. The reports needed were generated and printed on paper. Once the reports were completed, the tapes were placed in storage.

### **Relevance of the archive**

The U.S. Constitution mandates an official counting of the population every 10 years. While the official purpose of the Constitution’s mandate of a population count is to reapportion congressional districts, the Census also provides a statistical history of the nation and its people and is an economic asset and tool of inestimable value. The Census shows not only where people live at the time of the count, but also their educational levels, income, and other vital data.

In 1890, data were recorded with the aid of Hollerith’s punch cards. The 1950 Census of Population used a UNIVAC computer for tabulation of data. For the 1960 Census, the first to use the mail for collection of data, the Census Bureau and the National Bureau of Standards developed FOSDIC (film optical sensing device for input to computers), which was used until the 2000 Census. The surveys, completed by filling in dots opposite the appropriate answers, were photographed onto microfilm. FOSDIC read the dots and transferred the data to tape for computer input.

### **What happened to it?**

An internal 1963 Bureau of the Census technical memorandum listed tape files produced in connection with the 1960 Census of Population and Housing that the Bureau was retaining in “permanent data storage”.

Following consultation with staff of the National Archives in 1975, the Census Bureau created a plan to provide for the adequate retention of the 1960 data. They would retain 132 of the industry compatible tapes and would copy the tape files on 1,273 of the III-A tapes onto industry compatible tapes.

During 1975 and 1976, a member of the National Archives’ Machine Readable Archives Division reviewed the micro-aggregation or derivative files that the Bureau of the Census had preserved from the 1960 Census on the II-A tapes. This review identified seven series of low-level micro-aggregations as having long-term value. The seven series resided on 642 of the II-A tapes which the Census Bureau agreed to migrate onto more modern storage media, at least modern by 1975

industry standards. But by this time, the UNIVAC II-A tape drives were obsolete, and thus the preservation of these tapes presented a major engineering challenge.

### **Characteristics of the recovery**

Despite the challenge, the Census staff managed to find some old tape drives still in use that could read the tapes. These old drives were installed on a computer system that also had newer drives installed, so a tape conversion seemed simple. By 1979, the Census Bureau had successfully copied 640 of the 642 II-A tapes onto newer format tapes. The two tapes that were not copied were, in fact, missing. The missing tapes had 7,488 records, or about 0.5 percent of the total of approximately 1.5 million records on all II-A tapes that had been identified as having long-term value. Of the 640 tapes that were located, only 1,575 records (or less than 0.2 percent of the total number of valuable records on II-A tapes) could not be copied because of deterioration.

Hence a small volume of records from the 1960 Census was lost, and this occurred because of inadequate inventory control and because of the physical deterioration of a minuscule number of records. The bottom line is that 99.3 percent of the 1960 micro-aggregation data was saved on modern tape formats and can be read today.

The findings of the original study had already been published on paper in the 1960s, and that paper is preserved. The original 1960 U.S. Census documents were also recorded on microfilm, and all that microfilm is still in good condition, locked up at the National Archives. In compliance with U.S. laws, the complete 1960 U.S. Census documents on microfilm will be released to the public in the year 2032, 72 years after the original enumeration.

### **Cost of the recovery**

It has been impossible to find information about the overall cost of the recovery.

### **Conclusions**

The successful recovery of 99.3% of the micro-aggregation data marks a positive outcome which was made possible due to the availability of hardware similar to that used when the archive was originally created. The data which it was not possible to recovery was stored on tapes which were lost or had deteriorated.

<b>Archive</b>	Combat Air Activities File (CACTA)
<b>Purpose</b>	Database about combat air missions in Vietnam
<b>Institution in charge</b>	National Archives and Records Administration

### **Contents**

The Combat Air Activities File (CACTA) contains data from combat air missions in Southeast Asia, involving all the U.S. military services and the air force of South Vietnam. The “file” consists of 32 data files, each of which records information on the missions in a two-month period. The records describe air missions that flew between October 1965 and December 1970. No known data exists for November 1967. In 1971, the CACTA file was superseded by the Southeast Asia Data Base (SEADAB). Each CACTA record includes mission name and date; function and location of the mission; type, number, and identification of aircraft; results of the mission, including loss and damage data about aircraft and crew; and free-text comments. Records also contain data about bombing runs in Vietnam, including the ordnance used and the geographic coordinates where the ordnance was dropped.

### *Technologies used*

The Combat Air Activities File (CACTA), as several of the Vietnam-era data files held in the Center for Electronic Records of the National Archives, were created by the Department of Defense using an early data base management system called the National Military Command System Information Processing System 360 Formatted File System, commonly known as NIPS. The data structure of NIPS files is hierarchical in that each data record is composed of fixed, non-repeating data with one level of subordinating data. Each record is of varying length and is usually organized into the following sets of data elements: a Control Set, in which a unique record identifier is found, such as operations report number; a Fixed Set, containing non-repetitive data; and one or more types of Periodic Sets. Each type of Periodic Set may occur one or more times. In addition, NIPS files can include Variable Sets that appear only when data is present.

### **Relevance of the archive**

During the War in Vietnam, the US heavily bombed portions of the Ho Chi Min Trail in eastern Laos – over 2 million tons of bombs. As much as 30 % of the ordnance dropped did not explode. More than three decades later, farmers ploughing fields and children playing in bamboo thickets accidentally caused the ordnance to explode, which has killed or maimed more than 10,000 people.

An internationally funded program to identify potential unexploded ordnance is using targets, map coordinates, and type of ordnance for U.S. bombing runs from two Vietnam air combat databases that cover the years 1965–1975, which the National Archives accessioned in 1976-77. In Laos the data from the bombing runs were entered into a geographic information system that plots the target coordinates to the bombing database. Maps were then printed and used in the field by bomb technicians who use handheld GPS units to pinpoint the coordinates on the ground via navigational satellites.

### **What happened to it?**

The initial effort was unsuccessful because the geographic coordinates were flawed. The reason for these anomalies is that the data initially were created and used in a report generator system called the National Information Processing System (otherwise known as NIPS). NIPS was developed for the Department of Defense. By the time the records came to the National Archives they were legacy records and the staff on the Machine Readable Archives Division began a process of “deNIPSing” the records, that is moving them out of NIPS by reformatting them to a flat-file non-proprietary format in standard EBCDIC. The “de-NIPSeD” files are no longer dependent on the NIPS software

with which they were created. Instead, as flat files, users can process and manipulate them using widely-available software applications.

### **Characteristics of the recovery**

For 25 years it was believed that the “deNIPSeD” files were trustworthy reformatted records. However, the data anomalies found in the Combat Activities File raised a question. It now appears that at the time the “deNIPSeD” occurred, the documentation accompanying the data file either was incomplete or perhaps even missing, because the geographic coordinates, which were encoded in binary (binary angular measurement, a form of “packing data”) in order to conserve space, were incorrectly treated as 7-bit ASCII in each data field. Consequently, all of the geographic coordinates were wrong. Once this problem was identified, a process was begun to correctly convert these data fields.

### **Cost of the recovery**

It has been impossible to find information about the overall cost of the recovery.

### **Considerazioni finali**

The primary issue here is that migration introduced data anomalies into the records, which were not recognized for almost 25 years. These data anomalies could be corrected because NARA had retained the original NIPS data (that is the original 1 and 0s) and periodically migrated them to new media so that they remained “alive” or readable. The lesson to be learned here is the advantage of keeping the original bit stream alive.

<b>Archive</b>	Database of the <i>Consorzio Neapolis</i> (IBM and FIAT) - 1986
<b>Purpose</b>	Overall exploitation of environmental and artistic resources in the area around Mount Vesuvius
<b>Institution in charge</b>	<i>Soprintendenza Archeologica di Pompei</i>

### Contents

During the project, which lasted for two years (1987-1989), cost 36 billion Italian lire (financed as part of the *Giacimenti Culturali* cultural projects funding) and was carried out by a workgroup which consisted of as many as 110 people, the following activities were completed:

- The existent cartography of the area (that of ancient Pompeii above all) was converted into digital form and aerial photographs were taken
- A survey of archaeological emergencies in the area was made
- A survey of the archaeological monuments in the whole Pompeii complex was made
- A survey of the cultural heritage, of old town centres and the environmental heritage of the area was made
- Frescoes and mosaics found in the Pompeii archaeological complex were scientifically catalogued
- A number of excavation diaries of Pompeii (dating from the period 1862-1980) were digitised, including records and images with indications of the period to which each document relates and its current location
- A list was compiled of the period pictures relating to the Pompeii archaeological complex and of diaries kept by travellers in the 17<sup>th</sup> and 18<sup>th</sup> centuries

### Products developed

- *Base technical map*, in 1:5 000 scale, of the area, made from aerial photographs and their related recreation and storing in digital form.
- A number of *thematic maps*, in 1:25 000 scale (these also stored in digital form) and close ups in 1:5 000 scale of specific areas of interest for the project.
- Creation of a *photomap*, in 1:500 scale, made after *ad hoc* flights in 1987 over the archaeological complex of Pompeii, as well as an aerial photograph relief and digital storage of the same area in 1:500 scale. Both the photomap and the technical map are available in digital form.
- 1:1 000 scale map of the ancient city of Pompeii, divided into *regiones* and *insulae*, made by reworking Van der Poel's map – with the help of the physical recognition of around twenty archaeologists who corrected and completed the map by adding levels above and below that of the ground level and by surveying all the internal structures of archaeological interest (ovens, grinding wheels, guttering systems and so on).
- A complete address list of the buildings in Pompeii with information on the names of the houses (including the different names they have had in different historical periods), on their use and the types of buildings. This makes it possible to analyse the characteristics of Pompeii from a number of different aspects: commercial, leisure, distribution of public and religious buildings.
- Scientific cataloguing of the frescoes and mosaics of the Pompeii archaeological complex using the norms established by the ICCD, and their insertion into a *database*, giving each a biunivocal reference code which refers to the surface of the wall (or floor) which holds (or used to hold before its removal) the fresco (or mosaic), as well as to the record for the scientific catalogue and to the related documentation of the record itself, thus obtaining a clear indication of the relation between the object and the place where it was found.

- Creation of a *data management programme* which allowed – through the use of a specific language simplified for the user, questions and correlations going back to searches both of relational and free forms.
- Adoption, with the cataloguing record, of colour digital images, due to their stability over time with respect to traditional photographic images.
- Acquisition of both excavation diaries and period images (watercolours, drawings) into the database of discoveries.
- Representation of the spatial context both through the maps of the archaeological monument container of finds and through plans of the *insula* and the *regiones* of the area, by the creation of a plan of the entire archaeological complex.

For the entire project the capacity of the stored data to be reproduced has been taken account of, in order that the contents may be used in other seats of research and study. The data have therefore been stored on optical disks and magnetic tapes.

#### *Technologies used*

The IT system created worked on a mainframe of medium power, which made the simultaneous and concurrent use of different IT systems possible. The operating system made it possible for several users to use the system at the same time and housed a relational database system which managed the single and centralised database of the whole system.

Workstations were connected to the mainframe which carried out some of the elaboration phases synchronous to the central system locally. These workstations permitted the users to insert new data (records, images, digital maps) and also to find information easily.

#### **Relevance of the archive**

This was the first project in the cultural heritage sector in which an integrated centre for documentation was created, and also one that was extremely innovative from the technological point of view, in that it was based on a relational database which gathered “catalographic” information (information and images), maps and excavation diaries. The integrated system of use through all information contained in the database was the first example – and is still one of the rare cases – of cultural heritage databases which have made possible an integrated database of all the information relative to a certain site.

#### **What happened to it?**

The IT system created during the course of the project was officially handed over to the *Soprintendenza Archeologica di Pompei*, fully working, in 1989. In the same year, the computer in which all data and applications were stored, located in the Data Processing Centre in Boscoreale, was shut down due to a lack of financial and human resources by the Cultural and Environmental Heritage Ministry.

The system was reactivated, not without difficulty, in the period between October 1989 and May 1990, to allow the production of the touring exhibition *Rediscovering Pompeii*, organised by the Cultural and Environmental Heritage Ministry. In 1999, when the *Soprintendenza Archeologica di Pompei* decided to look into the possibility of recovering the information gathered by the Progetto Neapolis, they had to take into account the reactivating of the system. Even if this was possible, it would have required an unreasonable financial input.

For this reason a project to recover the data has been initiated, provisionally agreed to and approved by the AIPA, which has involved some of the people who had contributed the development of the original Progetto Neapolis. The recovery has been possible thanks to the economic resources of the *Soprintendenza* and was completed in 2000.

The *Soprintendenza* is currently involved in maintenance, update and enlargement of the existing database.

### **Characteristics of the recovery**

The recovery project has dealt mainly with the database which is the most important and valuable outcome of the Progetto Neapolis. As far as concerns the programmes for managing and accessing data, it has not been possible, given the technological changes which have taken place in the ten years since the system was shut off, to recover them. The company in charge of the recovery has given the *Soprintendenza* three new applications (for the ICCD cataloguing of the RA records and for the management of the excavation diaries) for the database management and a software application to view the information on the database using web techniques. As far as the database is concerned, at the end of the recovery project, the following materials have been produced:

- 32 385 RA records
- 21 732 excavation diary records
- 432 CAT records
- 337 BAT records
- 7 093 images relating to the RA records
- 27 394 images of the excavation diaries
- 29 images connected to the CAT records in .jpg format
- 1:1 000 scale archaeological map of the ruins of Pompeii in ArcView format
- 1:25 000 map of the archaeological emergencies of the area around Mount Vesuvius in ArcView format

Added to the material produced, the following have also been given to the *Soprintendenza*:

- 1 CD-ROM containing
  - Backup of the maps in ArcView and Autodesk Mapguide format
  - Backup of the SQLServer database
  - Copies, in Word format, of all paper-based documents.
- 7 CD-ROMs containing copies of all .jpg pictures.
- Paper-based documentation:
  - “User’s Guide” and “Installation Guide” for the application of navigation into the database.
  - User’s guide and functional specifications of the application “Management of Cultural Heritage”.

A fundamental condition for the recovery of the Progetto Neapolis data has been the availability, within the company in charge of the recovery, of a mainframe system with features very similar to those of the now disused one in Boscoreale, complete with tape drive, which has made it possible to read 140 tapes which made up a large part of the Neapolis archive, as well as to recreate the original working environment.

### **Cost of the recovery**

The cost of the recovery has been estimated by the company in charge of the recovery to be around 200 000 euro.

## **Conclusions**

The positive outcome of the recovery of the memory of the Neapolis project was made possible by:

- 4) the economic resources made available by the *Soprintendenza*
- 5) the availability of a processing system with similar features to that used in the original project
- 6) the identification and availability of people who had collaborated with the original project preserving its “oral memory”.

The cost of the recovery of the database should be considered as being extremely low in comparison with the cost of the original investment.

<b>Archive</b>	BBC Domesday
<b>Purpose</b>	Database describing how Britain looked to the British in 1986
<b>Institution in charge</b>	BBC Enterprises

## **Contents**

The BBC Domesday Project was a landmark multimedia resource which was produced to celebrate the 900th anniversary of the original Domesday Book - the document, regarded as the first great work of a bureaucratic state, commissioned by King William "The Conqueror" to provide a record of the land he had conquered twenty years before in 1066.

BBC Domesday was produced by a collaboration between the BBC, Acorn, Philips and Logica and cost 2.5 million pounds. A special hardware setup which included an expanded BBC Master and a LVRom player allowed the user to view the resource.

### *Products developed*

There were two discs in the Domesday Project package: the Community Disc and the National Disc. The basic difference is that the Community Disc is map-based and shows Britain as seen by the people who live there while the National Disc is topic-based and provides an overview of Britain.

The Community Disc begins with a map of the UK which you can zoom in on and scroll around. School children contributed photographic and written information about their local area.

The National Disc collects together a wide variety of statistical data, images of the UK and its culture in an array of photosets, surrogate walks (an early form of virtual reality) and a substantial amount of moving video.

### *Technologies used*

BBC Domesday runs on a BBC Master computer with several additional pieces of unusual hardware. The BBC Master was an 8 bit computer which saw great popularity along with the "BBC Model B" in the UK and Australia in the mid-1980s. Inside the Master a SCSI card and co-processor board were added. A trackball was plugged in to the Master, along with a special kind of enhanced video disc player called an LVRom.

The BBC Domesday multimedia application was written in a language called BCPL, chosen in an attempt to give Domesday some cross platform compatibility but which was not widely adopted, despite influencing the development of languages like C and Java.

The Domesday application software provides a multimedia front end to the mass of data stored on the video discs. Advanced search routines (believed to be a forerunner of Muscat), intuitive navigation and the ability to overlay different kinds of information in one view, made Domesday radically new in terms of interface design and as a powerful tool for its users.

## **Relevance of the archive**

The BBC Domesday Project was one of the major interactive projects of its time, and it was undertaken on a scale not seen since. The two discs contained more than 250,000 place names, 25,000 maps, 50,000 pictures, 3,000 data sets, 60 minutes of moving pictures and an unknown number of words. Over a million people - school children and researchers from across the country - took part in this ambitious project, and it is said that it would take you over seven years to look at everything on the discs. It may have been a marketing failure but those who made it saw it as a triumph of interactive knowledge engineering.

## **What happened to it?**

In November 1986 the discs were published by BBC Enterprises. They were available as a package together with the necessary hardware for around £5,000, which made them a little expensive for many schools and over twice what the project believed they were to cost when it started. By this time the British Government had changed its policy from subsidising hardware in schools to subsidising software. The best efforts of Philips and Acorn and even the help of ESPRIT funding from the European Commission could pare the cost down no more. As a result, few computers were purchased, and only a handful are left in existence.

BBC Domesday relies on a complex combination of hardware and software which was designed and manufactured in 1986. Many elements of this original setup are not compatible with today's computers. A small number of individuals and educational institutions still have working BBC Domesday systems, but these are becoming very scarce. Although the BBC Master is a remarkably hardy computer, after over 15 years of use many of the crucial LVRAM players have reached the ends of their lifetimes. Put simply, the BBC Domesday system is rapidly approaching complete obsolescence.

Ironically, the 15-year-old digital version has become unreadable, while the ancient paper one is in fine condition in the Public Record Office, Kew, and still perfectly usable.

## **Characteristics of the recovery**

The CAMiLEON project – Creative Archiving at Michigan and Leeds Emulating the Old on the New – has spent three years developing strategies for digital preservation. The BBC Domesday project has been chosen as a proof of concept test case for the emulation strategies they have developed.

CAMiLEON obtained access to a semi-working Domesday system donated by the School of Geography at the University of Leeds. One of the first tasks to preserve BBC Domesday was to transfer the data files from the 12 inch laserdiscs to bytestreams accessible on modern hardware. A Linux PC could be connected to the laserdisc player using a SCSI cable, allowing the PC to read the text articles and database. Images, including still-frame video, were transferred to a PC using a standard video frame-grabber card at maximum resolution. These images were stored in an uncompressed format to avoid quality loss. In total, around 70GB of data was transferred per side of each laserdisc.

The next step was to develop software that emulates the adapted BBC Master computer and the laserdisc player on which the original BBC Domesday system ran. The CAMiLEON team has successfully emulated BBC Domesday with its "DomesEm" demonstrator system, which included the co-processor, SCSI communication, and the many functions of the laserdisc player. Starting with the "BeebEm" BBC Micro emulator developed by David Gilbert and Richard Gellman, CAMiLEON developed the software to support the additional functionality required to run BBC Domesday. The emulation work was presented to an invited audience of digital preservation experts and members of the original BBC Domesday team on the 2nd December 2002.

## **Cost of the recovery**

It has been impossible to find information about the overall cost of the recovery. However, it must be stressed that the recovery is part of a bigger project aiming at the development of strategies for digital preservation

## **Conclusions**

A typical feature of the recovery – as well as the availability of an old Domesday system, albeit a semi-functional one – was the development of a software which emulated the original system and which could make up the future of methodologies and techniques of recovery of endangered archives.

<b>Archive</b>	East Germany's " <i>Kaderdatenspeicher</i> "
<b>Purpose</b>	Personal files database of East Germany's party functionaries
<b>Institution in charge</b>	German Federal Archives ( <i>Bundesarchiv</i> )

### Contents

The *Kaderdatenspeicher* contains personal data on 331,980 staff members (in 1989) of all former East German government agencies, excluding those of the Ministry of State Security, the Ministry for National Defense, and the Ministry of the Interior.

### *Technologies used*

In general, and also in the case of the *Kaderdatenspeicher*, East German computer centers used mainframe systems for the processing of large data compilations. The data holdings usually had hierarchical file structures that were not very complicated. The hardware and software used were copies of and variations on Western models, naturally with different names. For instance, the so-called ESER-mainframe systems in East Germany were copies of widespread mainframes.

As storage media, primarily 9-track tapes were used. Many of them had only a density of 800 bpi. Owing to production problems, these tapes bearing the East German trade marks ORWO or PYRAL proved to be in very poor condition. Glue and abrasion had to be removed from the tapes before they could be read. Sometimes, layers of the tape separated after the first reading because of insufficient binder.

The labeling of the tapes followed the international standards, with hardly any variation. Similar to Western mainframe applications, EBCDIC was used as code. The Russian code DKOI (in the former GDR also called "ESER-Code"), which in translation means Binary Code for Information Interchange, could also be found in East German data files. DKOI is very similar to EBCDIC and is basically an enlargement of EBCDIC with a few variations.

### Relevance of the archive

The files of the *Kaderdatenspeicher* not only provide insight into the political and professional career of officials, but they also contain information on their parents. There were several copies of the *Kaderdatenspeicher*, of which the only one that still exists is the one acquired by the Federal Archives. At least in one case, there is sufficient evidence that one copy of the *Kaderdatenspeicher* had been deliberately deleted shortly before the German unification in order to protect cadre members. The considerable value of this holding provided an incentive for the Federal Archives to invest quite heavily into the reconstruction of its documentation.

### What happened to it?

In October 1990, after 45 years of separation and different institutional traditions, the former East German Central State Archives were merged with the West German Federal Archives. When suddenly former East German government agencies and institutions were closed down, not only their paper records, but also a considerable number of machine-readable data holdings had to be secured or rescued from possible destruction. Very often, however, as in the case of the *Kaderdatenspeicher*, the documentation of these electronic records proved to be incomplete or even totally missing, therefore different approaches had to be taken to identify and verify data file structures and to reconstruct missing documentation.

After first copying the tapes of the *Kaderdatenspeicher*, the volume labels, the headers, and the first blocks of data of each file were printed out. The volume labels and headers followed widespread schemes, so it was easy to comprehend them. From these data, information on the content of each tape and an initial idea of the different generations and applications of the *Kaderdatenspeicher* could be obtained.

However, one typical problem already became apparent at this early stage. In the few lines of the volume label and headers, three different ways were used to express the date. Furthermore, the datasets of the *Kaderdatenspeicher* showed that only the full name, the Personal Identification Number, the address, and the agency were written in plain EBCDIC. All the other data fields were coded by numerical values, represented as binary figures.

### **Characteristics of the recovery**

Without a precise description of the data file structure, there was no way to understand the meaning of the data. Therefore, it became clear that as much information on the *Kaderdatenspeicher* as possible was needed. The orders and commissions to create and process the *Kaderdatenspeicher* came from the Council of Ministers. The vertical files of this office had been added to the collections of the Federal Archives in Potsdam after unification. After searching these holdings for references to the *Kaderdatenspeicher* project, a series of records that contains descriptions of the *Kaderdatenspeicher* and reports from the data processing center with a lot of substantial information could be found.

These paper records provided information on the content, purpose, history, and development of the *Kaderdatenspeicher* project, in particular:

- who planned the *Kaderdatenspeicher* and who gave the orders,
- which agencies cooperated,
- what the different aims and purposes of the *Kaderdatenspeicher* were,
- what information was contained in the *Kaderdatenspeicher*,
- how information was collected,
- which versions and updates of the *Kaderdatenspeicher* existed and which computer centers processed and stored them,
- who had access to which portions of the information contained in the *Kaderdatenspeicher*,
- and how information was used.

The reports to the Council of Ministers also contained information on the data file structure and codebooks. The *Kaderdatenspeicher* consists of annual compilations, so-called “generations” of data files for the year 1980 and for each year from 1985 to 1989, as well as of extracts for various purposes. Almost all of these data files have at least a slightly different structure. Nevertheless, the data file structures of all generations of the *Kaderdatenspeicher* could be found. Much information could be inferred from so-called “address tables”, which represent the record layout of a specific file.

The data files of the *Kaderdatenspeicher* were closely linked with the so-called staff databases of ministries and separate government branches, containing personal data of staff members. All the data of the *Kaderdatenspeicher* were originally collected from these staff databases. The Federal Archives has been successful in acquiring a relatively comprehensive and complete documentation of the staff database of the Ministry of Public Education. Therefore, additional information on the record layout and the data file structure of the *Kaderdatenspeicher* could be derived from the documentation of the staff database of this ministry.

In order to obtain background information, archivists have also made contacts with former employees of East German data processing centers who had created or worked with the data holdings which were acquired. In rare and difficult cases, for instance, when compression algorithms were used which could not be deciphered, programmers from former East German data processing centers were even hired as consultants.

### **Cost of the recovery**

The cost of the recovery, made possible thanks to a joint effort of the Federal Archives, the University of Jena and the Center for Historical Social Research in Cologne, has been estimated by the Direction of the *Bundesarchiv* at something between 150,000 and 200,000 EUR, two thirds of it being invested by the Federal Archives.

### **Conclusions**

The positive outcome of the recovery of the *Kaderdatenspeicher* was made possible due to two fundamental conditions: the discovery of paper-based documentation which, even though not directly descriptive of the archive, made it possible to identify the structure of the data files and therefore decipher their meaning, and the identification and collaboration of former employees of East German data processing centres who contributed, with their oral memory, to the deciphering of otherwise incomprehensible compression algorithms.

<b>Archive</b>	Websites of the city of Antwerp
<b>Purpose</b>	Preservation of the legacy websites of the city of Antwerp (4 versions: 1995-2001)
<b>Institution in charge</b>	City Archives of Antwerp; DAVID-project

## Contents

In 1995 Antwerp had its first website, called DMA, which contained general information about the City (council, agencies, events, agenda, news, private websites of citizens, etc.) and was developed by a huge team of web designers, webmasters at every agency of the city administration and volunteers. Several functions were available to the users: search, links to other websites, downloadable software for web developers, chat café, games corner, newsgroups, etc.

## *Technologies used*

HTML, GIF and JPEG images, CGI scripts, Java scripts, Apache webserver.

## Relevance of the archive

Antwerp was the first Belgian city which had a website, called *Digitale Metropool Antwerpen* (DMA). Pioneer websites of American and West-European cities such as *Digitale Stad Amsterdam* served as source of inspiration for the development of DMA. Version 1.0 (non-image and image) went online on June 11th 1995. Version 2.0 (non-frame and frame) had already been launched halfway through December 1995. Both versions have an historical value which justifies their preservation.

## What happened to it?

The legacy websites were preserved on backup tapes by the IT-staff of the city administration and transferred to the Antwerp City Archives in the fall of 2001.

However, Telepolis, the IT centre of the City of Antwerp, didn't have the necessary hardware and software configuration to restore the tapes to a hard disk. The backups had been made in a vendor dependant backup format on a SPARC computer with Solaris and an old Unix version as operating system. Telepolis wasn't using this computer system and backup software anymore. The files on the tape were also recorded according to the file system of the computer with which the backups had once been made, and a similar computer wasn't running anymore. On top of this, there was no documentation available. Therefore, the possibility of reconstructing the first versions of DMA was highly doubtful. In particular, there were problems with the old backup configuration of versions 1.5 and 2 (it was not available any more), and with the backup tape of version 3, which had been overwritten.

## Characteristics of the recovery

The IT staff managed to recreate the old backup configuration and to restore the content of the obsolete backup tapes on CD-ROM. As there was no written documentation, the main reason that made the recovery possible was the help of people who remembered non-written procedures such as passwords and old programming languages, and of people who were familiar with the old configuration. Besides, the recovery of versions 1.5, 2 and 4 was successful because there were still original data on the tapes, whereas that of version 3 was unsuccessful, because the source data were not available anymore.

The CDs were then transferred to the city archives and verified. In general, the files were 99 % platform independent. The only exception to this were the internal, absolute links and the counter which was based on a CGI script. The counter was not a problem, since it was going not to be active

in the archived version. Instead, the links within the recovered versions did cause some concerns. For many internal links, absolute paths had been used. When watching the website offline on CD, a lot of error messages or boxes with a red cross instead of a missing image were displayed. The amount of broken links was verified with the aid of “Linkbot”, a link checker. The control of the first version revealed 2509 broken links, whereas the website had a total of more than 64000 links. The programme indicated the web pages with broken links and the single links which didn’t refer to a specific file. The actual correction of the links had to be done manually.

A number of downloads in outdated file formats were also migrated to more recent file formats or more suitable archiving formats. A number of texts which were only available in WordPerfect 5.1, for example, were transferred to XML-files. After the conversion, the links to these documents were adjusted.

### **Cost of the recovery**

The cost of the recovery has been estimated by the Antwerp City Archives at three person/months, distributed as follows.

- IT staff: 8 weeks of work (3 weeks to recreate the old backup configuration; 4 weeks to adjust the internal hyperlinks; 1 week to create the CDs, the archivable versions, and to collect the metadata)
- Archivist: 4 weeks of work to develop the procedure, adjust the hyperlinks, make snapshots, and collect metadata.

### **Conclusions**

The positive outcome of the recovery – only partial due to the unavailability of one of the four versions due to the fact that the tape on which it had been stored had been overwritten – was made possible mainly due to the contribution of people who remembered non-written procedures and were familiar with the old configuration.