

La memoria perduta della scienza così si cancellano i dati digitali

ELENA DUSI

Riusciamo a leggere il codice di Hammurabi ma non i dati raccolti da Viking sulla Luna. La Nasa infatti ha perso la capacità di decifrare una parte dei bit inviati dalla sua sonda nel 1976. Né l'agenzia spaziale americana soffre da sola per il problema dell'invecchiamento di software e supporti digitali. Per l'80% delle ricerche condotte nel 1991 i dati sono andati persi. L'ha calcolato un gruppo di biologi dell'università della British Columbia, a Vancouver, che ha preso in esame 516 articoli scientifici pubblicati tra il 1991 e il 2011 su piante e animali. Studio per studio, i ricercatori hanno cercato di ripescare i dati su cui le pubblicazioni erano basate. E si sono ritrovati a mani vuote 4 volte su 5.

Se la scienza va avanti salendo sulle spalle dei giganti, le basi del passato sembrano ormai poggiate sull'argilla. Qualcuno l'ha chiamato il "medio evo digitale". Lo stesso fenomeno che rende illeggibili foto e video di famiglia di una manciata di anni fa sta erodendo i pilastri della scienza, sempre più fondata su grandi esperimenti e capace di raddoppiare la mole di dati prodotti ogni due anni. «Accade spesso che un esperimento sia irripetibile. Lasciare che i dati siano conservati

solo dai suoi autori vuol dire perderli nel giro di poco tempo, quindi buttare via risorse e frenare i progressi della scienza» è il commento sconcolato di Timothy Vines, il coordinatore del sondaggio.

Per non finire nello stesso buco nero di Viking, il Cern di Ginevra ha dedicato un palazzo di due pia-

I ricercatori di Vancouver: la rivista che pubblica un articolo si impegna a preservare le informazioni

ni alla memoria dei suoi esperimenti. In 80mila dischi e 50mila nastri sono preservati 100 petabyte di informazioni: mezzo secolo di fisica europea. Responsabile del Servizio conservazione dati è Alberto Pace, ingegnere proveniente dal Politecnico di Milano. «Il materiale più affidabile è il nastro magnetico. Dura a lungo, costa poco, non è soggetto a banchi devastanti e non consuma energia. Solo i dati che vengono consultati correntemente dagli scienziati sono su disco. Nel nostro centro di calcolo le procedure di controllo, aggiornamento e copiatura vanno avanti a ciclo continuo. Ogni singolo dati viene

riprodotto in media ogni quattro anni. Tutto il materiale è presente in almeno due copie, conservate in luoghi diversi».

Ma conservare i dati è una cosa. Riuscire a leggerli a distanza di tempo un'altra. E sempre al Cern Paolo Giubellino — che dirige l'e-

sperimento Alice — ricorda: «Qualche anno fa abbiamo avuto la necessità di ripescare i dati di un vecchio esperimento americano. Non c'è stato altro da fare che ricostruire da zero il computer con cui i risultati erano stati raccolti in prima battuta».

Nel centro di ricerche di Ginevra i dati raccolti dall'acceleratore di particelle in un anno occuperebbero una pila di cd alta 20 chilometri. Per la loro ricerca, invece, gli scienziati canadesi hanno usato studi basati su banche dati molto più semplici. A differenza dei grandi esperimenti su scala continentale, gli studi condotti da piccoli gruppi di scienziati affidano la conservazione delle misurazioni a uno dei loro autori.

«Ci aspettavamo delle difficoltà nel reperire informazioni di 50 anni fa. Ma scoprire che la quasi totalità dei risultati di 20 anni fa è andata persa è stata una brutta sorpresa» prosegue Vines. Ogni anno che passa dalla pubblicazione di un articolo aumenta del 7% la chance che dalla casella mail dell'autore arrivi la risposta "indirizzo sconosciuto". E cresce

del 17% la probabilità che lo scienziato abbia perso i dati, il

Il Cern di Ginevra ha dedicato un palazzo alla conservazione dei suoi esperimenti: 80mila dischi e 50mila nastri

software necessario a leggerli non sia diventato obsoleto o, in alcuni casi, perfino che l'autore rifiuti di condividere i suoi risultati.

In alcuni casi i ricercatori interpellati hanno risposto a Vines e colleghi che il loro computer era stato rubato, o perfino che si trovava in soffitta a casa dei genitori. Ma la maggior parte delle volte il materiale era conservato su floppy disk o altri supporti non più leggibili dai calcolatori di oggi. Per ovviare al problema dell'obsolescenza digitale, i ricercatori di Vancouver chiedono che la rivista che pubblica l'articolo si faccia carico della preservazione dei dati, di cui in genere l'articolo pubblicato è un semplice sunto. «Solo così — conclude Vines — eviteremo di perdere materiale a ritmi tali da restare a bocca aperta, e potremo permettere alla futura generazione di scienziati di avvalersi dei risultati del passato».

© RIPRODUZIONE RISERVATA

Si decifra un codice antichissimo, ma non si possono più leggere gli esperimenti recenti. Accade persino alla Nasa. E molte volte i supporti sono inutilizzabili



Cos'è il "Medio Evo digitale"

L'impossibilità di leggere i dati conservati in formato digitale solo pochi anni prima

Le cause

- I software più avanzati non riescono a leggere i dati memorizzati con programmi vecchi
- I supporti della memoria diventano obsoleti (è il caso dei floppy disk)

Un fiume di dati

1 milione di gigabyte i dati prodotti ogni secondo dall'acceleratore di particelle del Cern Lhc

- I dati prodotti dalle scienze raddoppiano ogni due anni

I supporti la durata media

5 anni hard disk
30 anni nastro magnetico

La Nasa

- Ha perso alcuni dati registrati dalla sonda Viking sulla Luna (1976) perché non esistevano più programmi per leggerli né i programmatori erano più in attività

Il Cern

100 petabyte Archiviati nel suo deposito @

- Alcune centinaia di megabyte vengono persi ogni anno a causa della rottura di hard disk non ancora salvati su nastro

Lo studio di oggi

- Ha preso in esame 516 articoli pubblicati tra il 1991 e il 2011 su riviste specializzate di biologia
- Si è andati a cercare i dati originali su cui gli articoli erano basati

80% dei dati è diventato irreperibile a 20 anni dalla pubblicazione

Le cause

- Le e-mail degli autori non sono più valide
- I supporti di memorizzazione sono diventati illeggibili

-17% ogni anno Le probabilità di trovare i dati originali di una ricerca

