

Software Heritage è un archivio mondiale dei codici sorgente, che conserva e rende pubblico il sapere immateriale alla base del mondo digitale

Custodire la storia dei codici del software

C'è un pezzo della nostra cultura che rischia ogni giorno di dissolversi nel niente. È il sapere immateriale con cui abbiamo costruito il *software* alla base delle nostre vite digitali, miliardi di codici sorgente sviluppati nel tempo da anonimi programmatori.

Finora a nessuno era venuto in mente di metterlo da parte, di conservarlo, catalogarlo e renderlo pubblico, come si fa in un museo con le ossa dei dinosauri. Poi qualcuno si è rimbeccato le maniche e ha cominciato a farlo. Ha messo in piedi un progetto basato su un'idea di *open science* su scala globale e ha convinto l'UNESCO a riconoscere il software tra i patrimoni culturali dell'umanità. Lui si chiama Roberto Di Cosmo e la sua grande idea è Software Heritage.

Come è nata l'idea di costruire un archivio globale del software?

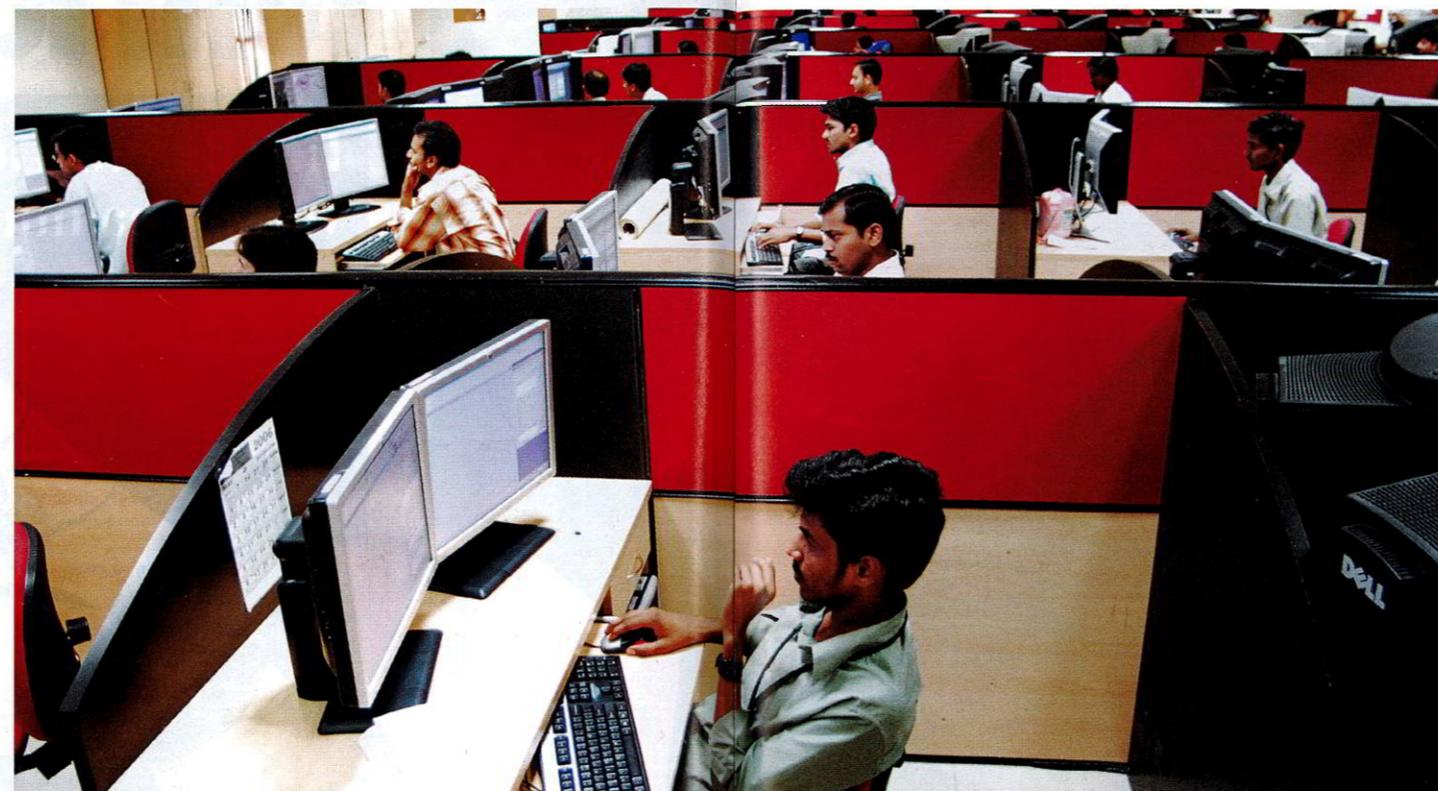
Il codice sorgente del software riveste un'importanza critica nella società sia come patrimonio culturale sia come infrastruttura tecnica fondamentale. A un certo punto ci siamo resi conto che stava diventando necessario preservarlo. L'idea si è concretizzata all'inizio del 2015, per iniziativa mia e di Stefano Zacchiroli, oggi professore all'Institut Telecom di Parigi, e con il supporto fondamentale di INRIA, l'istituto nazionale di ricerca in scienze informatiche e matematica applicata francese. L'obiettivo era creare un archivio universale che conservasse il codice sorgente disponibile, e garantisse l'accessibilità e la riutilizzabilità del software per le future generazioni. Ecco Software Heritage. Poi il progetto si è evoluto ed è diventato un punto di riferimento internazionale: ha espanso la sua rete di collaborazioni e ha aumentato significativamente il volume di codice archiviato.

Quali sono i numeri di Software Heritage?

A oggi, Software Heritage ha archiviato oltre 18 miliardi di *file* provenienti da oltre 270 milioni di progetti software, raccolti da centinaia di piattaforme collaborative utilizzate da circa 70 milioni di sviluppatori. Sono numeri impressionanti, che riflettono l'immensa diversità e ricchezza del patrimonio software mondiale e che mostrano l'importanza dell'archivio anche come risorsa per la ricerca e lo sviluppo delle scienze computazionali.

Quali sono le collaborazioni che la sostengono?

Software Heritage gode del sostegno di una rete di collaborazioni ampia e diversificata, che include istituzioni accademiche (in Italia l'Università di Pisa e quella di Bologna, la Scuola norma-



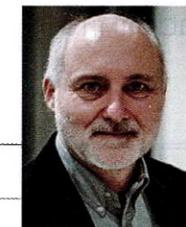
le superiore di Pisa e l'ENEA), aziende del settore tecnologico, organizzazioni non governative ed enti governativi. E poi c'è l'UNESCO, che riconosce il software come parte integrante dell'eredità culturale globale e supporta Software Heritage per il suo ruolo unico nel preservarlo.

Noi promotori del progetto siamo di origine italiana, e la sede di Software Heritage è in Francia: probabilmente questo spiega la nostra sensibilità rispetto a un patrimonio storico, e anche la presenza di molti attori francesi e italiani. Ma l'ambizione è di coinvolgere sempre più paesi anche al di fuori dell'Europa.

E i singoli sviluppatori possono contribuire a Software Heritage?

Certo, e in molti modi. Possono contribuire alla costruzione dell'infrastruttura dell'archivio, che è interamente software libero, e aiutare a identificare e archiviare progetti software anco-

di Silvia Bencivelli



CHI È ROBERTO DI COSMO

Roberto Di Cosmo, laurea alla Scuola normale superiore di Pisa e dottorato all'Università di Pisa, ha insegnato all'Ecole Normale Supérieure di Parigi e oggi è professore ordinario di informatica all'Université Paris Cité. Le sue ricerche spaziano dall'informatica teorica

al software libero, al quale ha contribuito dal 1998 con pubblicazioni e progetti di grande impatto come Mancoosi (MANaging the COmplexity of the Open Source Infrastructure) e con la direzione di IRILL (Initiative pour la Recherche et l'Innovation

sur le Logiciel Libre). La sua opera più rilevante è Software Heritage (<https://archive.softwareheritage.org>), realizzata con INRIA (Institut National de Recherche en Sciences et Technologies du Numérique) e UNESCO.

Ingegneri del software a Hyderabad, in India. Il codice sorgente del software ha un'importanza critica nella società.

fre comunque molto basse rispetto a quello che Software Heritage dà alla società. Ma anche rispetto al costo di una qualsiasi infrastruttura di ricerca.

Ci può fare esempi di software «conservato» da Software Heritage, che altrimenti si sarebbe perso?

Esempi emblematici sono i codici sorgente di programmi storici come quelli dell'Apollo 11 o di Amaya, il primo navigatore Web del W3C. O le prime linee di codice scritte sulla CEP, la Calcolatrice elettronica pisana (il primo calcolatore scientifico costruito in Italia), e del primo sistema di musica elettronica. Noi garantiamo che le future generazioni possano studiarli, visto che il codice sorgente, come dice Harold Abelson [il fondatore di Creative Commons, N.d.R.], è fatto in primo luogo perché gli esseri umani possano comprenderlo, e poi perché le macchine possano eseguirlo.

Che cosa cambia per Software Heritage con l'arrivo dell'intelligenza artificiale?

L'intelligenza artificiale (IA) cambia radicalmente il panorama del software, non solo nella creazione di nuovi codici ma anche nella modalità di conservazione e analisi del software esistente. Software Heritage in questo svolge un ruolo cruciale perché fornisce il miglior *dataset* disponibile al mondo per l'addestramento di algoritmi di IA. Lo dimostra il rilascio recente di StarCoder2, il più efficace modello aperto per la programmazione, addestrato con il nostro *database* e disponibile dal 20 febbraio su Hugging Face, la piattaforma *open source* dell'intelligenza artificiale, con la collaborazione di ServiceNow e Nvidia e del progetto BigCode. Ci aspettiamo che succeda anche con altri, per continuare lo sviluppo di modelli IA responsabili e trasparenti.

E da Software Heritage che cosa ci si aspetta per il futuro?

L'idea è diventare l'archivio mondiale di riferimento per tutto il codice sorgente, e per tutti gli usi: biblioteca di Alessandria del software, archivio di riferimento per la scienza aperta e la riproducibilità, referente unico per l'industria e grande telescopio per l'esplorazione della «galassia dello sviluppo del software». Con l'obiettivo ultimo di condurre l'umanità a produrre software di migliore qualità, più sicuro e più efficace per tutti.

Cortesia Roberto Di Cosmo (ritratto); A. Prabhakar Rao/The India Today Group via Getty Images (al centro)