

CONFE
RENZA
GARR
2026

LA FORMA DEL CAMBIAMENTO

Tra etica digitale
e radici condivise

SELECTED
PAPERS



CONFERENCE
GARR
2026

LA FORMA DEL CAMBIAMENTO

Tra etica digitale
e radici condivise

Pisa, 19-21 maggio 2026

SELECTED
PAPERS

 Consortium
GARR

Conferenza GARR 2026 - La forma del cambiamento
Tra etica digitale e radici condivise
Pisa, 19-21 maggio 2026

ISBN 978-88-946629-5-5

DOI 10.26314/GARR-Conf26-proceedings

Quest'opera è distribuita con Licenza Creative Commons Attribuzione 4.0 Internazionale (CC-BY).



Editore: Associazione Consortium GARR
Via dei Tizii, 6, 00185 Roma, Italia
www.garr.it

Curatori editoriali: Marta Mieli, Carlo Volpe
Progetto grafico: Carlo Volpe
Impaginazione: Marta Mieli

Prima stampa: Settembre 2026
Numero di copie: 350
Stampa: Tipografare
Via della Magliana, 1098, 00148 Roma (RM)

Tutti i materiali relativi alla Conferenza GARR 2026 sono disponibili all'indirizzo:
conf26.garr.it/it/

Indice

- 8 **Riprendersi le chiavi della casa digitale. Autonomia digitale, infrastrutture critiche e consapevolezza etica nell'era post-quantistica**
Massimo Carboni
- 13 **La sovranità digitale europea: programma, velleitarismo o riconoscimento di debolezza?**
Paolo Passaglia
- 17 **Il rilevamento dei contenuti generati dall'IA nelle istituzioni accademiche e di ricerca: limiti tecnici, implicazioni etiche e indirizzi operativi**
GianMarco Schiesaro
- 22 **Open Source Software Management at INGV**
Chiara Montagna, Stefano Cacciaguerra, Alessandro Pignatelli, Carlo Marcocci, Daniele Bailo, Elisa Trasatti, Francesco Martinelli, Giuseppe Messina, Mario Locati, Pasquale Cantiello, Placido Montalto, Roberto Tonini, Rodolfo Puglia, Rossana Paciello, Stefano Pintore, Tomaso Esposti Ongaro, Valentino Lauciani, Carmine Grasso, Claudia Fratianni
- 26 **ECHOTWIN: Un Ecosistema Digitale Integrato per il Continuum Edge-Cloud-HPC**
Arnaud Ceol, Alessandra di Benedictis, Lucia Floresta, Ugo Becciani, Davide Salomoni
- 31 **Abilitare l'ecosistema digitale con dati ai ready: il modello it4lia**
Donatella Sforzini, Arnaud Ceol
- 35 **From redundancy to governable continuity: energy-aware resilience and digital sovereignty for networks and datacenters**
Gianluca Mazzini
- 39 **Il Joint Lab per la Cybersecurity di FBK: governance Zero Trust, conformità NIS2 e Sovranità Digitale per gli enti di ricerca**
Leonardo Canello, Matteo Brosolo
- 43 **Transforming work practices through digital technologies: behavioral changes and cybersecurity challenges**
Mauro Sodani
- 48 **The INGV Centre for Computational Geosciences**
Tomaso Esposti Ongaro, Emanuele Casarotti, Giuseppe Di Stefano, Stefano Cacciaguerra, Paolo Perfetti, Marco Aliotta, Daniele Melini
- 52 **HPCC (High Performance Computing on Cloud): Workflow Orchestration Low-Code per l'Accesso alle Infrastrutture HPC on cloud**
Simone Giusepponi, Giulio D'amato, Andrea Spinelli, Marco Catillo, Luigi Bucci, Angelo Mariano, Francesco Buonocore, Sara Marchio, Giovanni Ponti, Massimo Celino

- 58 **The Paradox of the Synthetic Patient: AI, Data, and Theory in Translational Medicine**
Guido Gigante, Luisa Minghetti
- 63 **Automazione guidata della compliance nei progetti di ricerca: integrare DMP, GDPR e AI Act**
Leonardo Gaetano Mezzina
- 67 **SoBigData RI: Infrastructure, Governance, and FAIR Services for Social Data Science**
Michela Natilli, Roberta Savella, Laura Pollacci, Valerio Grossi, Roberto Trasarti
- 72 **Data governance challenges in multicentre biomedical research: the NeuroRehab4EU case**
Gaia Falorsi, Andrea Mannini, Federica Vannetti, Traian Popa, Andrea Serino
- 77 **Il nuovo portale Antenati e le nuove tecnologie: storie e dati prendono voce**
Marco Cavietti
- 82 **A distributed and AI-assisted framework for multimodal psycholinguistic tasks on desktop and mobile devices**
Ikram Massaoudi
- 87 **Web archiving "attivo" con IA: reingegnerizzazione assistita (LLM/Codex) e rilascio containerizzato di un portale di ricerca legacy**
Michele Fiaschi

Conferenza GARR 2026

La forma del cambiamento

Tra etica digitale e radici condivise

Pisa, 19-21 maggio 2026

Comitato di programma

Claudia Battista GARR

Elis Bertazzon GARR

Paolo Bolletta GARR

Massimo Carboni GARR

Raffaele Conte CNR

Giuseppe De Marco Dipartimento per la Trasformazione Digitale

Sara Di Giorgio GARR

Fabio Gadducci Università di Pisa

Gabriella Paolini GARR

Maria Chiara Pievatolo Università di Pisa

Lucio Pileggi Scuola Normale Superiore di Pisa

Giulio Selvaggi INGV

Mirella Serlorenzi Istituto Centrale per l'Archeologia

Stefano Suin Università di Pisa

Sabrina Tomassini GARR

Davide Vaghetti GARR

Chiara Veninata Istituto Centrale per gli Archivi

Simona Venuti GARR

Carlo Volpe GARR

Tutte le presentazioni e maggiori informazioni
sono disponibili sul sito dell'evento:
conf26.garr.it/it/

Web archiving “attivo” con IA: reingegnerizzazione assistita (LLM/Codex) e rilascio containerizzato di un portale di ricerca legacy

Michele Fiaschi

Scuola Normale Superiore

Abstract. Molte piattaforme di ricerca universitaria si fondano su database e logiche server-side; una loro semplice staticizzazione comporterebbe la perdita di indici, funzionalità di ricerca e relazioni tra entità, elementi che spesso costituiscono parte essenziale del loro valore scientifico. Il lavoro propone un web archiving “attivo”, che preserva il servizio nella forma originaria con il supporto di un agente di coding basato su LLM (Codex) supervisionato da un profilo non sviluppatore. Caso di studio è il portale La fortuna visiva di Pompei (PHP/MySQL legacy); il risultato è un rilascio Docker con runtime aggiornato, hardening documentato ed evidenze di regressione, ottenuto in circa sei ore: un metodo alla portata delle risorse ordinarie di università e PA.

Keywords. Web archiving, software preservation, LLM/Codex, legacy systems, containerization

Introduzione

Le piattaforme web accademiche, soprattutto nelle digital humanities, nascono spesso da progetti a finanziamento temporaneo, con stack non sostenuti nel lungo termine. Quando queste piattaforme diventano obsolete o insicure, le alternative sono di fatto tre: una copia “passiva”, che sacrifica ricerca, indici e relazioni; una reingegnerizzazione completa, con costi insostenibili per la PA; oppure lo spegnimento del servizio. Il web archiving classico mira a conservare e rendere accessibili nel lungo periodo risorse digitali pubblicate sul web, generalmente attraverso processi di crawling, conservazione in formato WARC e successivo replay delle risorse archiviate (ISO, 2017). Questo paradigma incontra però limiti con le applicazioni data-driven, in cui contenuti, indici e navigazione sono generati da query e relazioni in database. Nella loro tesi Isolani e Mezzina (2020) documentano progetti conclusi tra il 1999 e il 2013 con applicazioni obsolete ma ancora consultate e di valore, tra cui La fortuna visiva di Pompei, dove la copia statica non preserva l’elemento rilevante, l’interrogazione delle relazioni tra entità.

Si propone qui un approccio intermedio, il web archiving “attivo”: preservare non solo “la pagina” ma la fruibilità del servizio nella modalità di consultazione originale, con una reingegnerizzazione mirata, accelerata da un agente LLM e supervisionata da un tecnico.

1. Caso di studio e baseline tecnica

Il caso riguarda il portale La fortuna visiva di Pompei, raccolta di testimonianze testuali e visive sulla fortuna iconografica di Pompei, la cui consultazione dipende da una base dati e

da relazioni tra entità parte integrante del valore; l'interoperabilità (OAI-PMH) è parte del design originario. L'analisi del repository ha identificato tre macro-componenti: backend CMS/gestionale, frontend pubblico basato su PRADO e provider OAI-PMH. Le dipendenze sono legacy (PRADO incluso nel repo, niente Composer, librerie datate): l'aggiornamento selettivo è impraticabile e conviene riallineare il runtime contenendo il rischio. Lo strato dati è un DB MySQL 5.5 con 47 tabelle: indici, liste e schede correlate dipendono da query e relazioni non riducibili a pagine statiche. La codebase mostra pattern PHP 4/5 (API mysql_, hashing md5) e query concatenate, che suggeriscono un approccio per strati. L'obiettivo non è modernizzare la piattaforma, ma renderla eseguibile e pubblicabile in modo controllato, con hardening documentato, in un artefatto pronto per un sistemista con Docker.

2. Metodo: pipeline IA-first supervisionata

Il metodo mira a un artefatto di web archiving attivo: una versione eseguibile, riproducibile e documentata della piattaforma, distribuita come stack Docker aggiornato (PHP 8.x, MySQL 8.x) con bootstrap deterministico dei dati, verifiche di regressione e hardening documentato. Tre le domande guida: se un agente LLM possa portare a un rilascio riproducibile senza comprometterne controllabilità e sicurezza (RQ1); quali controlli minimi lo rendano affidabile in PA/ricerca (RQ2); quanto l'approccio "attivo" sia complementare a staticizzazione e astrazione dei dati (RQ3).

2.1 Ruoli e governance

L'agente IA ha svolto le attività a più alta intensità tecnica: lettura della codebase, individuazione degli hotspot, patch incrementali, script e documentazione. Il supervisore non era uno sviluppatore PHP né un sistemista, ma un progettista di servizi web con competenza architetturale e familiarità con Docker che ha definito i vincoli e verificato i risultati osservabili senza leggere ogni modifica al codice. Il metodo non richiede competenze tecniche specialistiche sui sistemi legacy, ma soprattutto capacità di coordinamento e gestione del processo. Questo lo rende più facilmente adottabile da università e Pubbliche Amministrazioni, dove tali profili sono già presenti, abbassando la soglia necessaria per replicare l'esperienza. Regola operativa: nessun intervento è accettato senza evidenze o test riproducibili.

2.2 Pipeline operativa e criteri di accettazione

La pipeline ha sei fasi incrementalì con logica di quality gate: fotografia tecnica (baseline); ambiente riproducibile (Docker); migrazione runtime e compatibilità; hardening di sicurezza; packaging e consegna "sistemista-ready"; raccolta delle evidenze. I controlli pass/fail comprendono lint PHP sui file modificati, smoke test HTTP su frontend, backend e OAI Identify, test di configurazione del runtime; indicatori qualitativi: paginazione, indici, allegati, coerenza OAI. Gli esiti convergono in un giudizio di pubblicabilità. Il percorso (oltre 40 step documentati) segue una disclosure controllata sui dettagli sensibili.

3. Implementazione e risultati

Il bootstrap automatico del database al primo avvio (DB, dump, schema OAI) ha reso testabili UI ed endpoint OAI Identify, il "laboratorio" per validare le patch. La migrazione a PHP 8.3 e MySQL 8 ha richiesto: gestione di null e array, paginazione stabile, ripristino di link e navigabilità delle schede correlate, accesso alle immagini via symlink e direttive Apache. Gran parte del lavoro consiste nel rendere esplicite e correggere assunzioni implicite del codice legacy, che risultano incompatibili con runtime più recenti e rigorosi.

Verifiche applicate a ogni ciclo di patch: il lint PHP ha sempre restituito PASS, gli smoke test HTTP hanno confermato stabilità degli endpoint e navigazione tra indici, schede e risorse. Il security assessment formalizza l'hardening su runtime PHP, Apache (header di sicurezza, mitigazioni upload-to-RCE) e override di produzione (DB non esposto, no-new-privileges, tmpfs), oltre a un hardening anti-SQL injection sui flussi esposti, con porzioni legacy non ancora su prepared statements. Le condizioni di pubblicabilità sono esplicite (TLS su reverse proxy, segreti robusti, monitoraggio), il perimetro dichiarato (nessun pentest completo). La pipeline si chiude con un pacchetto "sistemista-ready" (compose, template .env, guida operativa, checksum SHA-256): un oggetto di conservazione eseguibile e integro.

4. Discussione

Ciò che è stato preservato non è il runtime storico ma la funzione: navigare per indici, consultare schede e relazioni, accedere agli allegati, interrogare l'endpoint OAI. Un sistema data-driven non è riducibile alle sue pagine: uno snapshot non ne preserva le relazioni tra entità, in cui risiede il valore. Si accetta una trasformazione controllata del contesto esecutivo, documentata e verificabile: l'artefatto è un successore funzionale a condizioni esplicite, non una copia autentica. La risposta a RQ1 è affermativa, con la precisazione (RQ2) che i controlli minimi sono una condizione di funzionamento: il valore dell'agente non è accelerare la scrittura del codice, ma abbattere la barriera tra competenza architetturale e implementativa.

Rispetto a RQ3, tre strategie non esclusive: il web archiving attivo via Docker (S1) preserva la fruizione più completa al prezzo di una dipendenza infrastrutturale; la staticizzazione assistita riduce superficie d'attacco e costi ma sposta il carico sulla produzione; l'astrazione del dato rende il dataset citabile e riusabile (principi FAIR) indipendentemente dal runtime. Qui S1 è stato il percorso principale e la base delle altre due: il runtime funzionante ha consentito a Codex di estrarre dati e relazioni in forma sintetica, creando una versione statica e mantenendo la navigazione per indici (Puppeteer/Playwright, JSON locali) e un pacchetto pronto per la pubblicazione su Zenodo. La sperimentazione ha prodotto un risultato operativo in circa sei ore, evitando una reingegnerizzazione esterna stimata in mesi di lavoro, con costi difficilmente sostenibili e tali da compromettere il mantenimento della piattaforma online.

5. Conclusioni

Il lavoro mostra che un agente di coding basato su LLM, governato da un supervisore

non sviluppatore entro vincoli e quality gate verificabili, può produrre un rilascio Docker riproducibile che preserva la fruizione di un portale legacy data-driven. Il limite principale è la natura di single-case study: portali con workflow redazionali attivi, autenticazione complessa o dataset in crescita potrebbero porre sfide diverse. Direzioni future: l'esercizio reale con pentest formale e remediation assistita; l'adeguamento del frontend al design pubblico (Bootstrap Italia, AGID, accessibilità); la sperimentazione del metodo su più casi di studio, per individuare quali fasi della pipeline siano generalizzabili e replicabili in contesti applicativi diversi.

Riferimenti bibliografici

ISO (2017), Information and documentation - WARC file format, ISO 28500:2017, International Organization for Standardization, Ginevra.

Isolani A., Mezzina L. G. (2020), Gestione dei dati digitali della ricerca: linee guida per garantirne la produzione e conservazione sostenibile, tesi di Master SUM VIII, Politecnico di Milano.

Autore



Michele Fiaschi michele.fiaschi@sns.it

Michele Fiaschi è Responsabile del Servizio Sistemi Informativi della Scuola Normale Superiore dal 2013. Coordina progetti IT per didattica, ricerca e amministrazione, con focus su trasformazione digitale, dematerializzazione, sicurezza, protezione dei dati e accessibilità. Si occupa inoltre dell'adozione di piattaforme istituzionali e dell'applicazione di automazione e intelligenza artificiale in ambito universitario e PA.