

Una breve nota sulla costruzione di una base dati FAIR per la demografia storica

FRANCESCA TOSI, MARIO MARINO, SIMONE DI GIOVANNI

Alma Mater Studiorum - Università di Bologna

1. Introduzione: dati storici, *open science* e FAIRness della ricerca

La costruzione di basi di dati storiche richiede spesso un investimento rilevante in termini di raccolta, armonizzazione e verifica delle fonti. Questo lavoro preliminare, tuttavia, rimane di frequente poco visibile e scarsamente documentato. Ne derivano una limitata possibilità di confronto tra studi, una ridotta cumulatività della ricerca e una difficoltà nel valutare la solidità empirica delle analisi proposte.

Parallelamente, la corretta gestione dei dati secondo il paradigma FAIR (*Findable, Accessible, Interoperable, Reusable*) e i principi dell'*open science* e hanno assunto negli ultimi anni un ruolo centrale nel dibattito metodologico delle scienze sociali. In ambito storico-demografico, tuttavia, l'applicazione di tali principi solleva questioni specifiche, legate alla natura delle fonti, alla loro discontinuità nel tempo e alla difficoltà di rendere pienamente trasparenti e replicabili processi di ricerca fondati su materiali statistici storici.

Il database EcoPop_ER_Hist_Prov nasce in questo contesto e si colloca all'interno della famiglia di banche dati EcoPop_ER, sviluppata nell'ambito del progetto MEMOREC - Resilienza e Memoria Climatica in Emilia-Romagna (Di Giovanni, Scalone 2026). Tale progetto integra dati demografici, climatici e territoriali con l'obiettivo di documentare l'impatto delle condizioni ambientali sulla popolazione regionale nel breve e nel lungo periodo. EcoPop_ER_Hist_Prov ne rappresenta la componente storica, dedicata alla ricostruzione e alla diffusione di fonti statistiche demografiche relative all'Emilia-Romagna a partire dall'Unità d'Italia.

Se la versione contemporanea di EcoPop_ER è stata concepita come una banca dati georeferenziata per l'analisi spazio-temporale delle vulnerabilità sociodemografiche in relazione agli eventi climatici estremi, EcoPop_ER_Hist_Prov si propone come un'infrastruttura informativa per lo studio delle relazioni tra fattori ambientali e dinamiche demografiche nel lungo periodo, a partire da fonti statistiche storiche. In questa prospettiva, l'obiettivo del database non è la produzione immediata di indicatori operativi, bensì la messa a disposizione di una base informativa solida, trasparente e riutilizzabile, che consenta analisi comparative e cumulative nel tempo.

Lo sviluppo della banca dati EcoPop_ER_Hist_Prov è promosso congiuntamente dal progetto MEMOREC - Resilienza e Memoria Climatica in Emilia-Romagna, finanziato dall'Università di Bologna, e dal progetto PRIN 2022

‘Weather and Climate Vulnerability in Italian Demographic History’, finanziato dal Ministero dell’Università e della Ricerca nell’ambito del Piano Nazionale di Ripresa e Resilienza, nel quadro più ampio del programma NextGenerationEU dell’Unione Europea.

È opportuno precisare che EcoPop_ER_Hist_Prov non è stato utilizzato direttamente negli articoli empirici del presente numero. Il database viene invece proposto come una risorsa aperta e in larga misura ancora da esplorare, con l’obiettivo di favorire nuove analisi e confronti territoriali e temporali sulle dinamiche ambientali e demografiche nel breve e medio periodo.

2. EcoPop_ER_Hist_Prov: obiettivi, struttura del database e collocazione scientifica

Il database EcoPop_ER_Hist_Prov è progettato per fornire nuovi strumenti informativi per lo studio delle relazioni tra variabilità ambientale e dinamiche demografiche nella regione Emilia-Romagna. A tal fine, raccoglie e rende disponibili serie storiche mensili di eventi demografici a livello provinciale – matrimoni, nascite, decessi e nati morti, con disaggregazione per sesso – a partire dall’Unità d’Italia (Ministero dell’Agricoltura, Industria e Commercio 1862-1916, Ministero per il Lavoro e la Previdenza Sociale 1917, Ministero dell’Economia Nazionale 1918-1923).

Nella versione attualmente disponibile, la copertura temporale si estende dal 1862 al 1923 ed è strutturata in modo da consentire future estensioni cronologiche. Accanto alle variabili temporali *MESE* e *ANNO*, il database comprende, per ciascuna provincia emiliano-romagnola, le seguenti variabili demografiche:

- Numero di matrimoni registrati nel mese e anno indicati (*MATRIMONI_provincia*).
- Numero totale di nati registrati nel mese e anno indicati (*NATI_TOTALE_provincia*).
- Numero di nati maschi registrati nel mese e anno indicati (*NATI_MASCHI_provincia*).
- Numero di nate femmine registrate nel mese e anno indicati (*NATI_FEMMINE_provincia*).
- Numero totale di nati morti registrati nel mese e anno indicati (*NATIMORTI_TOTALE_provincia*).
- Numero di nati morti maschi registrati nel mese e anno indicati (*NATIMORTI_MASCHI_provincia*).
- Numero di nate morte femmine registrate nel mese e anno indicati (*NATIMORTI_FEMMINE_provincia*).
- Numero totale di decessi registrati nel mese e anno indicati (*MORTI_TOTALE_provincia*).
- Numero di decessi maschili registrati nel mese e anno indicati (*MORTI_MASCHI_provincia*).
- Numero di decessi femminili registrati nel mese e anno indicati (*MORTI_FEMMINE_provincia*).

Come frequentemente accade nella ricerca storica, le fonti utilizzate presentano discontinuità temporali, variazioni nella struttura delle rilevazioni e lacune infor-

mative che riflettono cambiamenti istituzionali o limiti della produzione statistica dell'epoca. La banca dati EcoPop_ER_Hist_Prov documenta in modo esplicito tali caratteristiche, mantenendo i valori mancanti come celle vuote e rinunciando deliberatamente a qualsiasi ricostruzione, interpolazione o stima *ex post*. Questa scelta metodologica risponde a un principio di fedeltà documentaria: EcoPop_ER_Hist_Prov restituisce le fonti per ciò che sono, rendendo visibili i limiti informativi della statistica storica e lasciando a chi conduce la ricerca la piena responsabilità interpretativa.

3. FAIR by design e pubblicazione dei dati

Fin dalle fasi iniziali, la pianificazione di EcoPop_ER_Hist_Prov è stata concepita secondo un approccio 'FAIR by design'. Il gruppo di ricerca ha riflettuto in modo esplicito sull'intero ciclo di vita dei dati. Dalla raccolta e organizzazione delle fonti, incluso lo studio sulle condizioni di riutilizzo, alla definizione della struttura della banca dati e della nomenclatura delle variabili, fino alla scelta del repository più adeguato alla disseminazione.

Tale riflessione si è concretizzata nella realizzazione di un Data Management Plan (DMP), che riassume le azioni adottate per gestire i dati nel modo più adeguato a garantirne la *FAIRness*.

Il database è stato pubblicato su Zenodo (Di Giovanni, Scalone 2026), piattaforma sviluppata dal CERN e ampiamente utilizzata dalla comunità scientifica per la conservazione e la diffusione dei dati di ricerca. Zenodo è un *trusted repository* che assegna identificatori persistenti (DOI), permette il versionamento – sia automatico che manuale – dei dataset e rende esplicite le condizioni di riutilizzo attraverso licenze standardizzate.

La scelta di Zenodo si inserisce in una strategia già adottata per le altre componenti del progetto MEMOREC e per la versione contemporanea di EcoPop_ER, rafforzando l'idea di una famiglia di banche dati accomunate da un'impostazione metodologica condivisa e orientate allo studio delle interazioni tra ambiente e popolazione su diverse scale temporali.

EcoPop_ER_Hist_Prov è attualmente disponibile in versione beta (v1, gennaio 2026), con DOI assegnato a livello di versione. Il deposito include, oltre ai dati in formato tabulare, un articolato pacchetto di metadati, concepito come parte integrante del database e come condizione essenziale per la sua riutilizzabilità. La documentazione acclusa ai metadati comprende:

- Un file di testo README, disponibile in italiano e in inglese, che descrive il contenuto del database e illustra le principali scelte metodologiche adottate nella sua costruzione.
- Un file excel contenente l'elenco delle fonti utilizzate per il reperimento dei dati mensili in serie storica a livello provinciale.
- Il *Data Management Plan*, con le informazioni relative alla gestione FAIR dei dati resi disponibili.

Ciascun file incluso nel pacchetto di metadati è concepito come documento dinamico, soggetto ad aggiornamenti continui al fine di riflettere l'evoluzione della banca dati.

4. Questioni legali e condivisione responsabile

Dal punto di vista giuridico, la costruzione di EcoPop_ER_Hist_Prov ha richiesto una riflessione sia sul diritto d'autore 'in entrata', legato al riutilizzo delle fonti, sia su quello 'in uscita', relativo alla condivisione del database finale. La natura storica e aggregata dei dati raccolti ha consentito di escludere criticità legate alla tutela della privacy, mentre il censimento delle fonti e delle rispettive condizioni di riutilizzo ha permesso di operare nel pieno rispetto della normativa sul diritto d'autore.

Il prodotto finale, pur configurandosi come una raccolta di dati funzionale a specifici obiettivi di ricerca, presenta anche le caratteristiche giuridiche di un database, in quanto risultato di un investimento significativo in termini di tempo, competenze e organizzazione delle informazioni. La scelta di adottare una licenza aperta CC BY 4.0 consente di valorizzare questo lavoro originale, promuovendo al contempo il riuso ampio e responsabile dei dati da parte della comunità scientifica.

5. Conclusioni: memoria climatica, *open science* e cumulatività della ricerca

Nel quadro del progetto MEMOREC, EcoPop_ER_Hist_Prov contribuisce alla ricostruzione della memoria storica dell'impatto delle condizioni climatiche e ambientali sulla popolazione, documentando – attraverso la costruzione e la diffusione di una base dati integrata, aperta e riutilizzabile – come le comunità dell'Emilia-Romagna abbiano reagito e si siano adattate nel tempo. In questo senso, la pubblicazione del database su Zenodo non costituisce un semplice atto conclusivo di disseminazione, ma rappresenta una componente strutturale dell'infrastruttura di ricerca secondo i principi FAIR della scienza aperta.

In tale prospettiva, EcoPop_ER_Hist_Prov è proposto come una risorsa messa a disposizione della comunità scientifica e della società civile, ancora in larga parte da esplorare, pensata per favorire nuove analisi sulle relazioni tra variabilità ambientale e dinamiche demografiche nel breve e medio periodo e per incoraggiare pratiche di ricerca più cumulative, trasparenti e replicabili.

Riferimenti bibliografici

- S. Di Giovanni, F. Scalone 2026, *EcoPop_ER_Hist_Prov (Beta version): Monthly provincial demographic time series for Emilia-Romagna, 1862–1923*, Zenodo. <https://doi.org/10.5281/zenodo.18378425>
- MEMOREC – Resilienza e Memoria Climatica in Emilia-Romagna, Progetto di ricerca interdisciplinare, Università di Bologna. Disponibile online: <https://site.unibo.it/memorec-resilienza-memoria-climatica-emilia-romagna/it>
- Ministero dell'Agricoltura, Industria e Commercio. Ufficio Centrale di Statistica 1862-1871, *Statistica del Regno d'Italia. Popolazione*, Firenze.
- Ministero dell'Agricoltura, Industria e Commercio. Ufficio Centrale di Statistica 1872-1874, *Statistica della popolazione. Movimento dello stato civile*, Roma.
- Ministero dell'Agricoltura, Industria e Commercio. Divisione di Statistica 1875-1876, *Statistica della popolazione. Movimento dello stato civile*, Roma.
- Ministero dell'Agricoltura, Industria e Commercio. Direzione di Statistica 1877-1878, *Statistica della popolazione. Movimento dello stato civile*, Roma.
- Ministero dell'Agricoltura, Industria e Commercio. Direzione della Statistica Generale 1879-1880, *Statistica della popolazione. Movimento dello stato civile*, Roma.

Ministero dell'Agricoltura, Industria e Commercio. Direzione Generale della Statistica 1881-1899, *Statistica della popolazione. Movimento dello stato civile*, Roma.

Ministero dell'Agricoltura, Industria e Commercio. Direzione Generale della Statistica 1900, *Statistica della popolazione. Movimento dello stato civile*, Roma.

Ministero dell'Agricoltura, Industria e Commercio. Direzione Generale della Statistica 1901-1910, *Movimento della popolazione secondo gli atti dello stato civile*, Roma.

Ministero dell'Agricoltura, Industria e Commercio. Direzione Generale della Statistica e del Lavoro. Ufficio Centrale di Statistica 1911-1916, *Movimento della popolazione secondo gli atti dello stato civile*, Roma.

Ministero per il Lavoro e la Previdenza Sociale. Ufficio Centrale di Statistica 1917, *Movimento della popolazione secondo gli atti dello stato civile*, Roma.

Ministero dell'Economia Nazionale. Direzione Generale di Statistica 1918-1923, *Movimento della popolazione secondo gli atti dello stato civile*, Roma.

Riassunto

Una breve nota sulla costruzione di una base dati FAIR per la demografia storica

Il contributo presenta EcoPop_ER_Hist_Prov, un database storico costruito secondo i principi FAIR (*Findable, Accessible, Interoperable, Reusable*) della ricerca per lo studio delle dinamiche demografiche in Emilia-Romagna. Il database raccoglie serie storiche mensili provinciali di matrimoni, nascite, decessi e natimorti, disaggregate per sesso, per il periodo 1862–1923. Il lavoro illustra le principali scelte metodologiche adottate, con particolare attenzione alla fedeltà alle fonti statistiche storiche, alla gestione esplicita delle lacune informative e alla disseminazione dei dati tramite il repository Zenodo, a supporto di pratiche di ricerca trasparenti e replicabili nella demografia storica.

Summary

A brief note on the construction of a FAIR database for historical demography

This paper presents EcoPop_ER_Hist_Prov, a historical database designed according to the FAIR principles (*Findable, Accessible, Interoperable, Reusable*) for the study of demographic dynamics in Emilia-Romagna. The database provides monthly provincial time series of marriages, births, deaths, and stillbirths, disaggregated by sex, covering the period 1862–1923. The paper outlines the main methodological choices, focusing on fidelity to historical statistical sources, explicit handling of missing information, and data dissemination through the Zenodo repository, supporting transparent and replicable research in historical demography.

Parole chiave

Open data; Principi FAIR; Serie storiche; EcoPop_ER_Hist_Prov.

Keywords

Open data; FAIR principles; Time series; EcoPop_ER_Hist_Prov.