INAIL

2026

I DATABASE IN TOSSICOLOGIA: FONTE INFORMATIVA DI IMPORTANZA CRESCENTE PER LA TUTELA DELLA SALUTE PUBBLICA E OCCUPAZIONALE

INTRODUZIONE

La tossicologia studia le risposte degli organismi agli xenobiotici (farmaci, inquinanti, altri fattori ambientali) e le possibili conseguenze patologiche, con l'obiettivo di tutelare salute umana, animale e ambientale. La valutazione della tossicità si è tradizionalmente basata su studi su organismi viventi e colture cellulari, metodi efficaci ma limitati da complessità biologiche, costi elevati, tempi lunghi e dall'esigenza di ridurre la sperimentazione animale. Oggi questi approcci sono affiancati da strategie innovative *in vivo*, *in vitro* e soprattutto *in silico* che sfruttano modelli computazionali avanzati e tecnologie ad alta processività per ottenere dati tossicologici più rapidi, predittivi e sostenibili. Gli studi *in vivo* permettono di valutare gli effetti sistemici delle sostanze su organismi animali o modelli complessi generando dati fenotipici, istopatologici e omici, raccolti in *database* come il *Chemical Effects in Biological Systems* (CEBS) e il *Toxicogenomics Project-Genomics Assisted Toxicity Evaluation system* (Open TG-GATEs). Le metodiche *in vitro*, basate su cellule, tessuti e organoidi per investigare meccanismi molecolari alla base di citotossicità, genotossicità o stress ossidativo, producono dati ad alta processività disponibili in appositi *database* quali *PubChem* e Open TG-GATEs. Gli approcci *in silico* utilizzano modelli matematici e bioinformatici per prevedere la tossicità a partire da strutture chimiche e reti biologiche, generando informazioni su *pathway* meccanicistici e simulazioni cinetiche avvalendosi di risorse come il *Comparative Toxicogenomics Database* (CTD) e la *Gene Ontology* (GO). La trasformazione digitale, sostenuta da *database* bioinformatici e tecnologie omiche, ha reso la tossicologia una disciplina sistemica e predittiva, integrando volumi di dati chimici, biologici e clinici, che migliorano interoperabilità e condivisione, oltre a ridurre l'uso della sperimentazione animale. Le tecniche omiche, quali genomica, trascrittomica, metabolomica, epigenomica forniscono informazioni dettagliate sulle risposte cellulari e molecolari agli agenti tossici, individuando alterazioni precoci e multilivello, integrandosi con i *database* tossicologici, infrastrutture digitali essenziali per raccogliere, standardizzare e rendere accessibili questi dati. Tra le principali piattaforme tossicologiche figurano CTD, CEBS e GO, che aggregano informazioni omiche e fenotipiche e supportano analisi meccanicistiche e predittive. L'uso combinato di diverse tipologie di dati biomedici, come parametri clinici (es. parametri vitali, biomarcatori), informazioni sui livelli di esposizione ambientale o occupazionale, dati epidemiologici riguardanti popolazioni (come l'incidenza di malattie) e dati omici relativi alle risposte geniche, proteiche e metaboliche, consente una valutazione del rischio molto più accurata per popolazione generale e lavoratori esposti, anche attraverso biomarcatori di esposizione, suscettibilità ed effetto precoce, come microRNA e polimorfismi genetici. Nei contesti occupazionali, dove sono frequenti esposizioni miste, i *database* integrano dati espositivi e biologici per rilevare alterazioni precoci e orientare la prevenzione. Per valorizzare appieno questi patrimoni informativi servono tecnologie avanzate come intelligenza artificiale (IA) e *machine learning* (ML): la prima per gestire l'eterogeneità dei dati e il secondo per analizzare *dataset* standardizzati. Le autorità regolatorie, in particolare nell'Unione Europea attraverso il regolamento REACH (Reg. CE 1907/2006) e l'Agenzia europea per le sostanze chimiche (ECHA), negli Stati Uniti l'Agenzia per la protezione dell'ambiente (EPA), incoraggiano l'uso di modelli *in silico* basati su *database* aperti, in linea con i principi delle 3R (*Replacement, Reduction, Refinement*). La crescente necessità di integrare fonti informative eterogenee ha portato allo sviluppo degli *Integrated Approaches to Testing and Assessment* (IATA). Questi approcci combinano proprietà chimico-fisiche, dati espositivi e informazioni cliniche, con risultati *in vitro*, *in vivo* e *in silico*, migliorando la decisione regolatoria in modo più rapido, efficiente ed etico.

I PRINCIPALI DATABASE IN TOSSICOLOGIA

I *database* tossicologici più diffusi e utilizzati contengono informazioni che riguardano gli effetti di xenobiotici e farmaci su cellule, tessuti, organi o sull'intero organismo, con dati derivanti da studi previsionali *in vitro*, *in vivo* o *in silico* concernenti diverse specie (Tabella 1). Allo scopo di raccogliere e analizzare dati provenienti da diverse fonti alcune Agenzie internazionali hanno catalogato in appositi *database* le informazioni tossicologiche sugli agenti chimici. È il caso dell'Agenzia americana per la protezione dell'ambiente (EPA), che ha istituito il *database* pubblico ACToR (*Aggregated Computational Toxicology Resource*), raccogliendo informazioni da agenzie americane ed europee, nonché da fonti accademiche. Tale piattaforma contiene dati derivanti da numerose fonti pubbliche relative all'identità, alla struttura, alle proprietà chimico-fisiche delle sostanze e alle fonti espositive, al fine di catalogare ed aggregare le informazioni per supportare la gestione e la valutazione del rischio ambientale. Un altro *database*, sviluppato dal *National Toxicology Program* (NTP), che archivia dati tossicologici, ambientali, molecolari e istopatologici su modelli animali e cellulari, provenienti da fonti accademiche, industriali e governative, è il *Chemical Effects in Biological Systems* (CEBS), che ha come principale obiettivo quello di promuovere la convergenza delle informazioni per nuovi studi integrati di metanalisi.
Il *Comparative Toxicogenomic Database* (CTD) è un "pro-

totipo" dei nuovi *database* pubblici, che si è evoluto quale solido punto di riferimento ed ecosistema digitale ed incorpora dati su sequenze e strutture di proteine e acidi nucleici, oltre che su composti tossici per la salute umana di origine ambientale. Il CTD si pone come piattaforma molto aggiornata nella quale i ricercatori possono, sulla base delle evidenze relative alle alterazioni dei *pathways* molecolari desunte dalla letteratura scientifica e da studi sperimentali, indagare ed individuare precocemente le relazioni di natura tossicologica tra esposizioni ambientali e *outcome* patologici, aspetto cruciale negli scenari complessi di esposizione a xenobiotici di diversa natura quali gli ambienti occupazionali. Un approccio avanzato per valutare in maniera ottimale i rischi associati alle esposizioni ambientali e occupazionali è dato dall'integrazione dei diversi *database* tossicologici più avanzati e aggiornati con tecniche statistiche e bioinformatiche, quali la *Gene Enrichment Analysis* (GEA), un *tool* del *Gene Ontology* (GO), che rappresenta a propria volta una fonte molto autorevole per informazioni su funzioni geniche, processi biologici e cellulari. Dato un *set* di geni, appartenenti a famiglie differenti, identificati con un vocabolario standardizzato e comparabile nel database GO, che siano, ad esempio, *up*-regolati in determinate condizioni espositive, la GEA può contribuire ad evidenziare associazioni statisticamente significative nell'ambito dei geni maggiormente coinvolti in processi quali alterazioni del ciclo cellulare e infiammazione, la cui espressione sia modificata dall'esposizione.

Tabella 1 **Esempi di alcuni database utili in tossicologia ambientale e occupazionale**

Database	Gestore/Ente	URL	Contenuto
Comparative Toxicogenomics Database (CTD)	*National Toxicology Program, National Institute of Environmental Health Sciences* (NIEHS)	https://ctdbase.org	Relazioni di natura tossicologica tra esposizioni ambientali e *outcome* patologici
Aggregated Computational Toxicology Resource (ACToR)	*US Environmental Protection Agency* (EPA)	https://19january2017snapshot.epa.gov/chemical-research/actor_.html	Dati strutturali su sostanze chimiche per la valutazione del rischio espositivo
Chemical Effects in Biological Systems (CEBS)	*National Toxicology Program, National Institute of Environmental Health Sciences* (NIEHS)	https://cebs.niehs.nih.gov/cebs/	Dati integrati di studi omici, molecolari, tossicologici, istopatologici
Gene Ontology (GO)	*Gene Ontology Consortium, National Human Genome Research Institute* (NHGRI)	https://geneontology.org/	Vocabolario per ontologie, processi biologici e cellulari da associare ad effetti tossici da esposizioni ambientali e occupazionali
PubChem	*National Center for Biotechnology Information* (NCBI)	https://pubchem.ncbi.nlm.nih.gov	Banca dati su struttura e bioattività delle sostanze chimiche, informazioni su tossicità, sicurezza e interazioni biologiche
Distributed Structure-Searchable Toxicity (DSSTox)	*US Environmental Protection Agency* (EPA)	https://www.epa.gov/comptox-tools/distributed-structure-searchable-toxicity-dsstox-database	*Database* chimico per il supporto alla modellistica applicata in tossicologia computazionale
Toxicogenomics Project-Genomics Assisted Toxicity Evaluation system (Open TG-GATEs)	*Toxicogenomics Informatics Project - National Institutes of Biomedical Innovation, Health, and Nutrition* (NIBN)	http://toxico.nibio.go.jp/english/index.html	Dati sperimentali di tossicogenomica per studi di tossicità molecolare e alterazioni fenotipiche

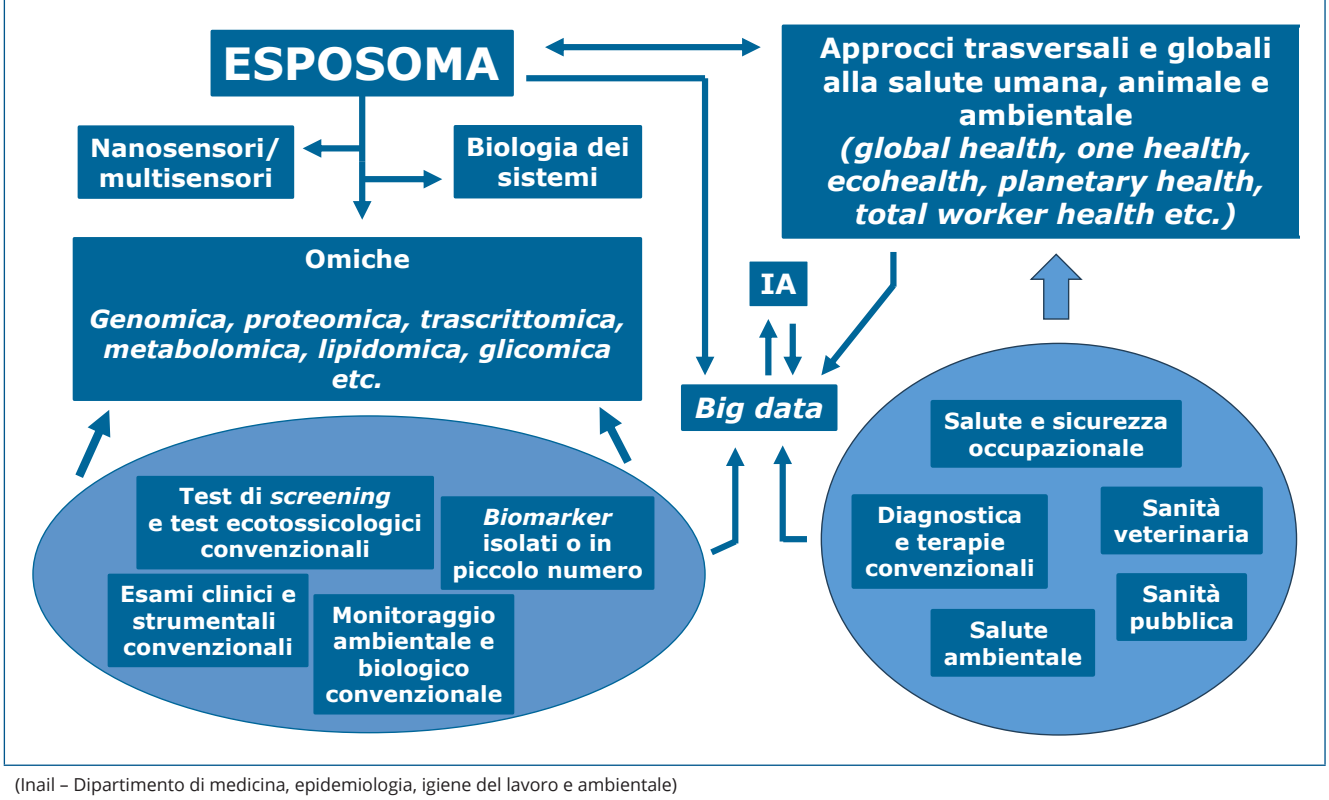
(Inail – Dipartimento di medicina, epidemiologia, igiene del lavoro e ambientale)

L'IMPATTO DEI DATABASE IN TOSSICOLOGIA

Negli ultimi decenni il rapido sviluppo nel campo della tossicogenomica e delle tecnologie computazionali ha reso disponibili enormi volumi di dati biomedici derivanti da parametri clinici, esposizioni, misurazioni biologiche, studi epidemiologici e profili molecolari, rendendo indispensabile l'allestimento di appositi *database* che hanno trasformato le modalità di valutazione degli effetti tossici delle sostanze potenzialmente dannose per la salute umana e per l'ambiente. Parallelamente, la salute e la sicurezza dei lavoratori stanno vivendo un rapido cambio di paradigma: ambienti di lavoro più complessi, influenzati dal cambiamento climatico e da esposizioni multiple, richiedono approcci integrati. Discipline prima separate convergono in modelli come *one health e total worker health*, mentre le analisi tradi-

zionali lasciano spazio a metodiche omiche, al concetto di esposoma e all'uso di *big data* e intelligenza artificiale. Questo processo coinvolge competenze multidisciplinari, dall'ingegneria alla sociologia, fino alle tecnologie avanzate. La figura 1 rappresenta schematicamente questo cambiamento. I *database* hanno consentito di digitalizzare e integrare un'enorme quantità di dati eterogenei, rendendo possibile una modellazione predittiva più mirata e contribuendo a ridurre l'uso di test animali. Hanno inoltre migliorato la prevenzione in ambito occupazionale e ambientale grazie a strumenti decisionali basati su evidenze e potenziato la ricerca, garantendo maggiore interoperabilità, trasparenza e accesso ai dati. La raccolta attraverso l'integrazione di tecnologie multi-omiche di informazioni strutturate e funzionali in apposti *database* riguardanti le complesse relazioni tra geni, metaboliti e proteine ha consentito di ampliare le conoscenze sugli *adverse outcome pathway* (AOP) nei sistemi biologici. L'obiettivo primario dei *database* pubblici è quello di favorire l'accesso, la condivisione, la diffusione e la trasparenza di informazioni dettagliate, facilitando lo scambio tra ricercatori, professionisti del settore sanitario e cittadini, al fine di trasferire l'innovazione dei risultati in termini di tutela della salute pubblica e occupazionale. In tale ottica è pertanto fondamentale sviluppare interfacce *user friendly* per agevolare la consultazione e l'interpretazione dei dati da parte degli utenti. In tal modo viene incentivata la condivisione di dati di alta qualità, catalogati sulla base di una terminologia e di formati standardizzati, indispensabili per una corretta comparazione e quindi per ridurre gli errori interpretativi e la formulazione di conclusioni errate. Tuttavia, per ottimizzare le informazioni presenti nei *database* tossicologici è necessario avvalersi anche e soprattutto di supporti basati su metodologie computazionali predittive altamente performanti, quali IA e *machine learning*, che devono essere integrati con approcci sperimentali all'avanguardia in tossicologia, quali genomica, proteomica, metabolomica, epigenomica. L'IA è una tecnologia utile per processare in modo efficiente enormi quantità di dati eterogenei, quali i dati biomedici, per garantire qualità, accuratezza, coerenza e interoperabilità dei dati, individuare alterazioni precoci, supportare decisioni complesse negli approcci IATA. Il *machine learning* a sua volta risulta più performante per analizzare *dataset* tossicologici già strutturati e standardizzati e predire *endpoint* di tossicità. Il ruolo dei *database* non si limita tuttavia all'archiviazione e alla catalogazione dei dati ma, come in parte già rilevato, deve supportare i ricercatori e i professionisti del settore nella individuazione di alterazioni molecolari precoci attraverso specifici biomarcatori innovativi, quali i microRNA, e di suscettibilità individuale, quali i polimorfismi genetici. In tal modo è possibile inquadrare e quantificare più efficacemente effetti sub-clinici, soprattutto nel contesto delle esposizioni occupazionali multiple, in particolare ad agenti chimici e fisici, consentendo un intervento tempestivo ed evitando la successiva evoluzione verso manifestazioni di tipo clinico. Sono inoltre possibili sostanziali miglioramenti nella previsione e nella prevenzione di potenziali effetti a lungo termine e nell'individuazione degli individui (lavoratori e non) maggiormente vulnerabili all'azione, singola o combinata, di *noxae* specifiche.

Figura 1 **Rappresentazione semplificata del cambiamento di paradigma concettuale e metodologico nella tossicologia ambientale e occupazionale, nelle scienze della salute e dell'ambiente**

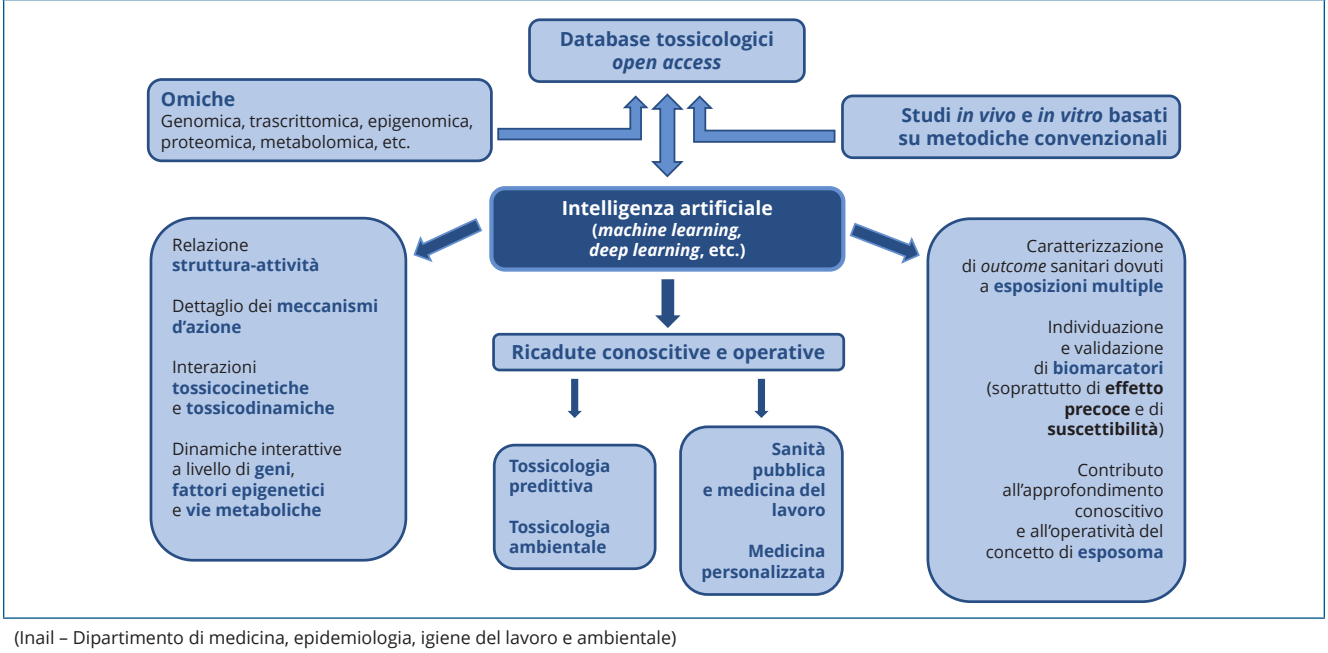


(Inail – Dipartimento di medicina, epidemiologia, igiene del lavoro e ambientale)

CONCLUSIONI

Negli ultimi tempi la tossicologia sta attraversando una vera e propria rivoluzione digitale, soprattutto grazie alla disponibilità di enormi volumi di dati provenienti da fonti di diversa natura (espositivi, molecolari, biochimici). L'applicazione dei *database* tossicologici di pubblico accesso rappresenta uno strumento indispensabile per catalogare, standardizzare, rendere accessibili e valorizzare pienamente le informazioni generate dall'integrazione tra le tecniche sperimentali omiche e le metodologie computazionali più avanzate. L'approccio sinergico tra *database*, tecniche omiche e computazione avanzata nella tossicologia contemporanea è ormai consolidato. È il solo che, consentendo di avviare a soluzione le problematiche legate a standardizzazione, eterogeneità, qualità e interoperabilità dei dati, rende possibile, facendo ampio ricorso anche alla predittività, gestire e superare molte delle attuali criticità nell'ambito della salute pubblica e occupazionale. Un prospetto sintetico di quanto illustrato nel presente *fact sheet* è riportato nella Figura 2.

Figura 2 **Schema di sintesi relativo ai database tossicologici open acces e all'impiego dei sistemi di intelligenza artificiale per finalità predittive e preventive nell'ambito dei rischi ambientali e occupazionali**



(Inail – Dipartimento di medicina, epidemiologia, igiene del lavoro e ambientale)

PER ULTERIORI INFORMAZIONI

Contatti: p.capone@inail.it; ca.grandi@inail.it

BIBLIOGRAFIA ESSENZIALE

Singh AV, Bansod G, Mahajan M, et al. Digital transformation in toxicology: Improving communication and efficiency in risk assessment. ACS Omega. 2023;8(24):21377-21390.

Davis AP, Wiegers TC, Sciaky D, et al. Comparative Toxicogenomics Database's 20th anniversary: update 2025. Nucleic Acids Res. 2025;53(D1):D1328-D1334.

Pawar G, Madden JC, Ebbrell D, et al. In silico toxicology data resources to support Read-Across and (Q)SAR. Front. Pharmacol. 2019;10:561.

PAROLE CHIAVE

Database; Tossicologia; Intelligenza artificiale

ISBN 978-88-7484-987-1 © 2026 Inail. La pubblicazione viene distribuita gratuitamente e ne è quindi vietata la vendita nonché la riproduzione con qualsiasi mezzo. È consentita solo la citazione con l'indicazione della fonte.
Stampato dalla Tipolitografia Inail di Milano • Edizione 2026 • **Autori: P. Capone, C. Grandi, P. Chiarella, R. Sisto** • Inail - Dipartimento di medicina, epidemiologia, igiene del lavoro e ambientale.
Progetto editoriale: Inail-Dimeila • Redazione editoriale e grafica: C.Branchi.

Inail - Dipartimento di medicina, epidemiologia, igiene del lavoro e ambientale - Direttore: Bruno Papaleo
Via Fontana Candida, 1 - 00078 Monte Porzio Catone (RM)
Tel. 06 941 81 405 - 404 • Fax 06 941 81 419 • dmil@inail.it