

WOBD: Web of Biological Data

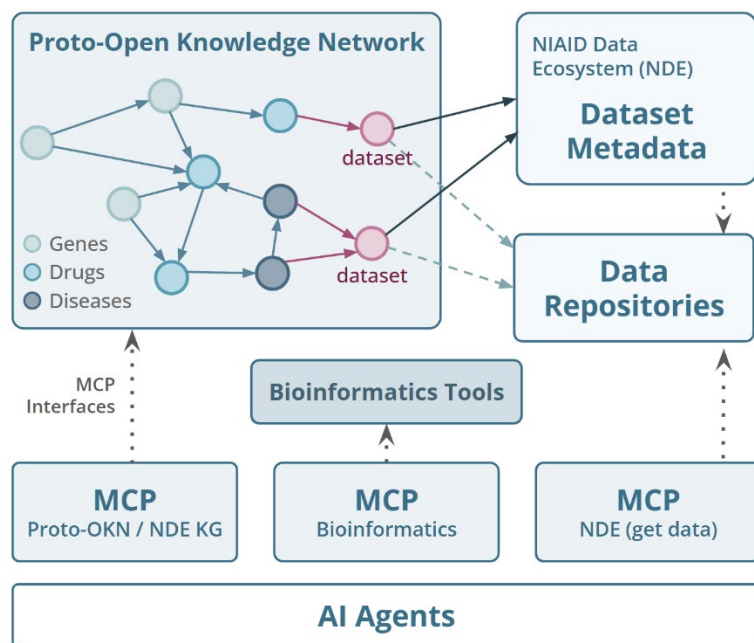
CONNECTING LIFE SCIENCE DATASETS ACROSS INSTITUTIONS



U.S. National
Science Foundation

THE VISION: In April 2025, the National Security Commission on Emerging Biotechnology recommended creating a Web of Biological Data (WOBD) as a “single point of entry for researchers to access high-quality data” to “advance U.S. leadership in biotechnology for national security and economic resilience”. WOBD spans the full range of biological data — from curated knowledge graphs (KGs) to raw -omics, imaging, sequence, and EHR data — and joining them unlocks opportunities for new biological discovery and synergistic data mining.

PROTOTYPE DEVELOPMENT: The NSF Open Knowledge Network (OKN) is an ideal launchpad for implementing this vision; we have prototyped a powerful, AI-ready OKN-WOBD that connects knowledge, metadata, and data across a broad spectrum of biological domains.



DID YOU KNOW?

Biological research data flourishes at the boundaries between datasets and disciplines, and data reuse maximizes our returns on investment in scientific research.

- **OKN-WOBD builds on OKN** — leveraging its cross-disciplinary infrastructure and deep bio-health KG coverage as the foundation.
- OKN-WOBD pulls in **harmonized dataset metadata** from dozens of repositories (indexed by the National Institute of Allergy and Infectious Diseases (NIAID) Data Ecosystem (NDE)) so **data stays at the source but becomes queryable** alongside the OKN's biological knowledge.
- OKN-WOBD exposes its knowledge through the **Model-Context Protocol (MCP)** which allows any AI tool (Claude, ChatGPT, etc.) to translate natural language questions into dynamic queries and analyses spanning all these layers.
- Continued expansion has **compounding value**: every new graph, repository, and metadata relationship widens the set of cross-domain questions WOBD can answer.

Available NOW at apps.okn.us/wobd

TWO ACCESS MODES:

- Guided queries — structured templates for common patterns (keyword search, differential gene expression, drug-disease traversal) that compile user inputs into SPARQL.
- AI assistant access — natural-language queries through the MCP, compatible with any MCP-enabled agent.

THREE DEMONSTRATED WORKFLOWS:

- Meta-analysis of gene expression in diabetic nephropathy
- Environmental sampling and molecular toxicology of PFAS "forever compounds"
- Identification of candidate host species for terpene biosynthesis